Computer Vision and Image Understanding 154 (2017) 1–15

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Computer vision for assistive technologies

M. Leo[a,*], G. Medioni[b], M. Trivedi[c], T. Kanade[d], G.M. Farinella[e]

[a] National Research Council - Institute of Applied Sciences and Intelligent Systems, Italy
[b] University of Southern California, USA
[c] University of California San Diego, USA
[d] Carnegie Mellon University, USA
[e] University of Catania, Italy

ARTICLE INFO

Article history:
Received 17 April 2015
Revised 1 September 2016
Accepted 2 September 2016
Available online 6 September 2016

Keywords:
Computer vision
Assistive technologies

ABSTRACT

In the last decades there has been a tremendous increase in demand for Assistive Technologies (AT) useful to overcome functional limitations of individuals and to improve their quality of life. As a consequence, different research papers addressing the development of assistive technologies have appeared into the literature pushing the need to organize and categorize them taking into account the application assistive aims. Several surveys address the categorization problem for works concerning a specific need, hence giving the overview on the state of the art technologies supporting the related function for the individual. Unfortunately, this "user-need oriented" way of categorization considers each technology as a whole and then a deep and critical explanation of the technical knowledge used to build the operative tasks as well as a discussion on their cross-contextual applicability is completely missing making thus existing surveys unlikely to be technically inspiring for functional improvements and to explore new technological frontiers. To overcome this critical drawback, in this paper an original "task oriented" way to categorize the state of the art of the AT works has been introduced: it relies on the split of the final assistive goals into tasks that are then used as pointers to the works in literature in which each of them has been used as a component. In particular this paper concentrates on a set of cross-application Computer Vision tasks that are set as the pivots to establish a categorization of the AT already used to assist some of the user's needs. For each task the paper analyzes the Computer Vision algorithms recently involved in the development of AT and, finally, it tries to catch a glimpse of the possible paths in the short and medium term that could allow a real improvement of the assistive outcomes. The potential impact on the assessment of AT considering users, medical, economical and social perspective is also addressed.

1. Introduction

The term "Assistive Technologies" (AT), in the broadest sense, refers to any set of scientific achievements (products, environmental modifications, services and processes) useful to overcome limitations and/or improve function for an individual (Cook and Polgar, 2014). More specifically AT aim at helping persons with disabilities or special educational/rehabilitation needs, to deal within their daily context and to achieve a better quality of life (Lancioni et al., 2012). In general there are two main application contexts in which AT are exploited: 1) the medical application context that tries to reduce (or rehabilitate) physical or cognitive impairments and 2) the social application context that operates on the surrounding environment focusing on social barriers and discriminations (Hersh and Johnson, 2010). In the last decades there has been a tremendous increase in demand for new technological solutions allowing an improvement of the quality of life, e.g., for elderly people or people with different abilities, as well as for people without disease but willing to increase their comfort. Many researchers, working in different fields, have applied their knowledge to build advanced technologies in order to meet the needs of diverse AT' application contexts (Bishop, 2014; Bourbakis et al., 2015; Dakopoulos and Bourbakis, 2010; Jo et al., 2014; Lancioni and Singh, 2014; Murphy and Darrah, 2015; Pawluk et al., 2015b; Velázquez, 2010; Vichitvanichphong et al., 2014; Vuong et al., 2015). As a consequence, works that deal (directly or indirectly) with the development of AT have proliferated in the literature and therefore the issue to properly organize them has emerged. This categorization task has been then carried out by grouping the works dealing with the same area of user needs: e.g. dementia (Vuong et al., 2015), visual diseases (Murphy and

* Corresponding author.
E-mail addresses: marco.leo@cnr.it (M. Leo), medioni@usc.edu (G. Medioni), mtrivedi@ucsd.edu (M. Trivedi), Takeo.Kanade@cs.cmu.edu (T. Kanade), gfarinella@dmi.unict.it (G.M. Farinella).

http://dx.doi.org/10.1016/j.cviu.2016.09.001
1077-3142/

Darrah, 2015), social orientation deficit (Bishop, 2014), activities of daily living for aged persons (Vichitvanichphong et al., 2014), severe/profound and multiple disabilities (Lancioni and Singh, 2014), physical and cognitive disabilities in children (Jo et al., 2014), smart environments (Velázquez, 2010). This type of categorization could be defined as "user-need oriented" and it is the foundation of all the existent surveys concerning AT: they are very interesting and useful for readers (medics, patients, companies, etc.) who address the problems related to a specific need, hence giving the overview on the state of the art technologies supporting the related function for the individual.

However in this common way of looking at the AT world, each technology is considered as a whole and a deep and critical explanation of the technical knowledge used to build the operative tasks is usually missing. As a consequence, the cross-contextual applicability of most of the underlying techniques and methodologies involved in each specific context is not considered and then the resulting surveys are unlikely to be technically inspiring for functional improvements and to explore new technological frontiers.

To overcome this critical drawback, in this paper a "task oriented" way to categorize the state of the art of the AT works has been introduced: it relies on the split of the final assistive goals into operative tasks that are then used as pointers to the works in literature in which each of them is used as a component.

The potential usefulness of this new way to categorize the state of the art of the AT works can be easily foreshadowed by considering that each technical task can be part of different AT dealing with quite different user's needs. For example object recognition from visual cues can be part of an assistive device for the indoor navigation for visual impaired persons, or part of an alternative communication interfaces, as well as a module in a social robot platform, and so on. Keeping this observation in mind, it follows that by collecting works dealing with the same task, regardless of the application context, it becomes easier to highlight the straightness and the open critical issues, allowing the reader to identify the research lines to be undertaken for an improvement of the existing assistive technologies.

Unfortunately, as already stated above, in literature there are no papers that use the "task oriented" approach and then this paper tries to partially fill this gap by using a set of cross-application Computer Vision tasks as the pivots to establish a categorization of the AT already used to assist some of the user's needs (Mental Function, Mobility, Sensory Substitution and Assisted Living) pointed by the World Health Organization. For each task the paper analyzes the Computer Vision algorithms recently involved in the development of AT and finally it tries to catch a glimpse of the possible paths in the short and medium term that could allow a real improvement of the assistive outcomes. The rest of the paper is organized as follows: in Section 2 the most common areas of users' needs are identified and, for each area, the list of the underlying relevant cross-application AT tasks involving Computer Vision algorithms are pointed-out. In Section 3, for each identified task, the analysis of the leading methods and techniques reported in literature to address it in any assistive context is given. Section 4 considers the assessment of AT considering users, medical, economical and social perspective. A discussion about open challenges is then supplied in Section 5 where a glimpse into the possible emerging new solutions is also given. Finally Section 6 concludes the paper.

2. Users' needs and related Computer Vision tasks

The range of human needs to be addressed by AT is quite large, however, according to the Flagship programme – Global Cooperation on Assistive Health Technology (GATE), developed by the world Health Organization (WHO) (http, 2015), most of them can be grouped into the following classes:

1. Mental functions
2. Personal mobility
3. Sensory functions
4. Daily living activities
5. Orthotics and prosthetics
6. Communication and skills training
7. Recreation and sports
8. Housing, work and environmental improvement

Among the above human needs, Computer Vision has been involved to build AT supporting the first four classes. The remaining classes have been otherwise addressed by scientific areas related to mechanics, electronics, computer graphics, signal processing, robotics, medicine, education, communication, material and building science. In the following, for each of the users' needs in which computer vision can be considered as a technological support, the most relevant research findings are indicated and the involved computer vision tasks are identified. Techniques and methodologies involved in each task will be then detailed and discussed in the next section. In Table 1 a scheme of the relationships between user's needs and the involved computer vision tasks is shown.

Mental functions impact the capacity of an individual to think, concentrate, react to emotions, formulate ideas, problem solve, reason, remember. Traditional assistive technologies supporting mental functions may include smart-watches, personal digital assistants and electronic calendars with alarm functions, voice guidance or graphic illustrations. Ordinary smartphones with internet access can also offer special user interfaces or applications to be used to support mental functions. Emerging assistive technology for supporting mental functions aim at expanding mental healthcare services by involving computer vision algorithms for the development of Socially Assistive Robot (SAR) (Gillespie et al., 2012; O'Neill and Gillespie, 2014) i.e. a robot that provides assistance to human users through social interaction (Mataric, 1999). SAR research has been carried out to study its impact on a different mental health concerns (e.g., dementia, depressed mood, autism spectrum disorder) on groups of patients belonging to different categories (e.g., children, elderly). Socially assistive robots have already served in a variety of clinically relevant roles: Companion, Therapeutic Play Partner, Coach/Instructor (Rabbitt et al., 2015). Socially assistive robots can take many different physical embodiments, including animal-like, machine-like, and human-like forms. The development of a SAR platform is a plethora of complex and multidisciplinary research areas and involve, among all, computer vision tasks such as visual self-localization and mapping (SLAM) in the environment (Meng et al., 2014), detection and tracking of objects (principally people), soft and hard biometrics (Carcagnì et al., 2015a), human action and activity recognition (Fasola and Mataric, 2013), head pose/gaze estimation (Cazzato et al., 2014), recognition of facial emotions/expressions (Medioni et al., 2007). The implementation of these functional modules allow the SAR platform to understand the surrounding environment, to react to its changes and to instantaneously adapt its behavior to the identity, mood and/or level of attention of the assisted person (Louie et al., 2014). Moreover, the robot is able to understand and control proxemics (the social use of environments) in order to employ mechanisms for communication which are analogous to those used by humans (Mead et al., 2013).

Aside from assistive mobile platforms, another interesting research area direct towards supporting mental functions, aims at analyzing human behaviors starting from visual data coming from environmental and wearable sensors. In particular, recognition of facial emotions/expressions and human action and activity recognition tasks are performed and evaluated over time through ad-

Table 1
User's needs and the relative computer vision tasks involved in the assistive technologies.

User's needs → CV tasks ↓	Mental functions	Personal mobility	Sensory functions	Daily living activities
Localization	x	x	x	x
SLAM	x	x	x	
Object detection	x	x	x	x
Object tracking	x	x	x	x
Human activity recognition	x	x		
Biometric	x	x	x	
Head pose estimation	x	x		
Gaze estimation	x			x
Image retrieval	x			x
OCR			x	

vanced machine learning and data mining methods. This area of research and technology development is sometime referred as "Behavioral Imaging" as analogy to the medical imaging research area that revolutionized internal medicine in the 20th century. Behavioral imaging provides efficient and objective measurements of behavior without the need of labor-intensive human observation. Moreover, it can be exploited to get real-time feedback that can immediately highlight the effectiveness of the ongoing assessment or therapy (Rehg et al., 2014).

Within this scientific field, different efforts are also devoted to the introduction of technological intervention tools to improve emotion processing skills through realistic and flexible stimuli. As an example, authors of Cassidy et al. (2016) described a method for the creation of near-videorealistic avatar, which can produce a video of a face uttering inputted text, with a variety of emotional tones. They demonstrated that general population adults can correctly recognize the emotions portrayed by the avatar. Adults with autism spectrum disorders were significantly less accurate than controls, but still above chance levels for inferring emotions from the avatar. Both groups were significantly more accurate when inferring sad emotions from the avatar compared to the original actress, and rated these expressions as significantly more preferred and realistic.

The use of detection and tracking of objects on mobile devices has been also used to superimpose digital content on physical objects creating augmented reality that can be exploited to guide on object usage (Damen et al., 2016), even through a robotic manipulator (Jiang et al., 2016), or to increase the selective and sustained attention of children with autism during object discrimination therapies and elicit more positive emotions (Escobedo et al., 2014). A research area which is extremely useful to support mental functions is the one referred with the name "First Person Vision" (Devyver et al., 2011a; Kanade and Hebert, 2012). The aim of First Person Vision is to sense the environment and the subject's activities and intents from a wearable sensor considering the point of view of the user (having this way readily available information about his head motion and gaze direction). Despite this kind of wearable computing has long history (Betancourt et al., 2015; Mann, 1997), it has become of more interest in the last years with the advancement of both hardware and software technologies. One of the main application of First Person Vision systems is related memory augmentation (Damen et al., 2014; Gurrin et al., 2014) (e.g., how to use an object?) and life logging (Furnari et al., 2015; 2016; Gurrin et al., 2014; Ortis et al., 2016). In this context the representation of semantic concepts appearing in visual lifelogs is a fundamental step since the huge amount of data collected by a user over time (personal big data) (Gurrin et al., 2014; Wang et al., 2016).

Personal mobility, concerns the individual's ability to move within an environment or between environments, as well as the ability to manipulate objects. Unfortunately, some persons have the mobility functions compromised by impaired body functions or structures. Some technologies support mobility problems either keeping under control the movements of injured body parts or making more effective rehabilitation of the body functions. Two examples of this kind of technologies are the Baclofen pumps (that allow a person to control his or her spasticity) and the robotic therapy devices (who allow people to reduce impairment through repetitive movement training). These kind of supports are referred as 'indirect' or 'therapeutic' technologies. Differently, assistive (or 'direct') technologies are exploited to support the individual mobility by introducing advanced devices that augment the mobility functionalities although they do not alter the impairment underlying the mobility loss. Cane, wheelchair or walkers and prosthetic limb are examples in this specific context (Cowan et al., 2012). Another area of research is devoted to the development of low-cost and easy-to-use personal mobility technologies to evaluate the nominal value and the variability of spatio-temporal gait parameters (such as length of steps, trunk orientation, gait events, etc.).

In this AT research fields, Computer Vision algorithms are involved in the innovative area of the prosthetic limb control where systems use image recognition to autonomously select the proper hand orientation, grasp shape and grasp size based on images of the object being manipulated (Došen et al., 2010). The visual controller performs the detection and tracking of the object, measures its geometric properties (size, shape) and automatically selects grasp type and aperture size appropriate for grasping the object. Computer Vision has been also exploited to build devices able to remotely measuring spatio-temporal gait parameters while providing visual video feedback. Measurements can be performed from data acquired either from static sensors or properly placed on board of mobile platforms able to follow the patient at a constant distance. In these case (people) detection and tracking can be effectively achieved by merging multisensorial data (coming from laser and infrared or omnidirectional cameras) or by using single or multiple cameras combined with the depth of the scene. In particular the latter scheme has been introduced with the aim to build an affordable mobile analysis platform for pathological walking assessment (Bonnet et al., 2015).

Mobility rehabilitation is often needed after stroke, surgery, or degenerative diseases. It has to be personalized for each patient and the tools can be easily calibrated through smart interfaces and virtual reality. Advanced rehabilitation systems can perform human action and activity recognition (Zeng et al., 2012) using vision-based techniques which ensure natural interaction experience. The integration of vision-based interfaces with thematic virtual environments allows the development of novel tools and services regarding rehabilitation activities (Avola et al., 2013). Robotics Agent Coachers can assist people with movement disorders during the execution of motor exercises: this operating mode generally increases the user participation by exploiting the game-like interac-

tions with the robotic agent and, at same time, allows up close human movement detection and measurement.

Computer vision techniques, have been also exploited in the development of intelligent wheelchairs (BrgidaMnica et al., 2015) making use of command interface that does not require the hands: the wheelchairs' software addresses problems like face detection, head pose estimation (Rivera et al., 2013) and/or recognition of facial emotions/expressions (Sobia et al., 2014) for generating signals for interfacing with the wheelchair. Vision based autonomous localization frameworks have been also used to make efficient navigation using a wheelchair: visual self-localization and mapping (SLAM) techniques have been used to calculate the wheelchair motion and then to face the doorway passing regardless of its starting position (Narayanan et al., 2014) as well as for the navigation along corridors (Narayanan et al., 2016; Pasteau et al., 2014).

Sensory functions: people with sensory impairments have reduced ability (or lack of ability) regarding vision, touch, and hearing senses. The effects of a sensory impairment can range from slight to complete loss of ability to use the senses with mild or severe impact on daily activities. Assistive technologies supporting sensory functions generally transform the characteristics of one sensory modality into stimuli of another sensory modality. Sometime loss of sensitivity in hands or fingers (e.g., due to peripheral neuropathy or other causes) can make it difficult or impossible to use a standard interfaces (as mouse, keyboard) and then alternative interfaces are needed. In this area Computer Vision techniques are mainly involved to support blind or people with reduced sight persons. The past few years have witnessed an exponential growth in the computing capabilities and onboard sensing capabilities of mobile phones making them an ideal candidate for building powerful and diverse applications in this context. Robust and efficient object detection and recognition (that include also text detection and recognition for text-to-speech AT) can help people with severe vision impairment to independently access unfamiliar indoor environments and avoid dangers. Applications for forgery detection and value identification of banknotes have been proposed in literature (Bruna et al., 2013). The great advancements on text recognition (Epshtein et al., 2010; Jaderberg et al., 2014) has allowed to bring the classic OCR to the commercialized level. Nowadays the text can be recognized by wearable systems which are able to read for the user in uncontrolled settings (i.e., there is no need to perform a scan of the document with an external hardware and calibrate settings). A commercial example of wearable device for this task is OrCam (OrCam, 2016).

Computer vision technologies have the potential to assist blind individuals to independently access, understand, and explore the environments (Tian et al., 2013). Visual Self-localization and mapping (SLAM) techniques have been also used to estimate the ground plane (Leung and Medioni, 2014): this can aid the blind in dynamic environments since it can be used by other high level navigation tasks such as obstacle avoidance, path planning, visual SLAM and self-localization frameworks. A wearable system composed mainly by an RGB-D camera for indoor navigation system able to complementing the white cane, has been presented in Lee and Medioni (2016). The system does not make use of a prior map or GPS information and exploit both sparse features and dense point clouds to get ego-motion estimation. Through a probabilistic mapping for path planning it is able to guide the visually impaired user from one location to another making. The system also allows to store and reloads maps allowing this way to expand coverage area of navigation. In (Chessa et al., 2016) a conceptual bio-inspired framework with identification and recognition capabilities able to provide early visual capabilities to impaired people has been described. It offers multi-functional aids based on Computer Vision for scene understanding and wayfinding. Also computer vision frameworks working in outdoor environments to assist people in road crossing action (Coughlan and Shen, 2013; Mascetti et al., 2016), finding the walkable path (Phung et al., 2016) or shopping (Chippendale et al., 2014) have been introduced. In this field, another issue to be faced by computer vision is related to soft and hard biometrics: the recognition of the persons in front of the assisted person (i.e. the possibility to have an accurate face recognition system) is an increasing demand from visual impaired users (Chaudhry and Chandra, 2015). A novel vision-based method to analyze the layout of a web page to facilitate access to web content for users with visual impairments was proposed in Cormier et al. (2016). Another useful application concerns with the design, development and evaluation of wearable mobile reading devices that rely on robust document image analysis in order to identify the structure of the document (Keefer and Bourbakis, 2014; Keefer et al., 2013; Koo and Cho, 2010). Virtual environments can allow the users to interact with different virtual structures and objects through auditory and haptic feedback and they can help people who are blind in training themselves in familiar and unfamiliar spaces (Lahav et al., 2015). In recent years, lip event analysis in videos has been extensively studied because of its attractable applications including lipreading, visual speaker identification, audioûvisual speech recognition (AVSR), human computer/robot interaction, facial expression analysis and so forth. In particular efficient lip event detection approaches have been introduced: they, without learning priors, aim not only to distinguish frames depicting visual speech from those depicting visual silence, but also to investigate the lip-dynamic states of mouth opening and closing (Liu et al., 2016). Finally techniques to improve navigation outcomes for prosthetic vision users have been also proposed. Prosthetic vision users often cannot interpret complex, uncontrolled scenes and then computer vision and image processing techniques can be used in some systems to improve functional outcomes (Horne et al., 2016).

Daily living activities: In the last years, several research projects have been focused on systems for socio-medical services with the aim of introducing ICT technologies to increase context user awareness and improve his comfort, safety or independence while performing daily activities. The main aspects influencing the perception of quality of life concern possible degeneration of health status, safety and social life. As a result the technologies in this area must provide the following main services: general health status monitoring, indoor and outdoor localization, domestic environment monitoring and remote control interfaces. Health and safety status monitoring are generally performed by wearable systems equipped with sensors and electronics modules. Computer vision-based solutions have been also recently employed for safety and health monitoring: they address object detection and tracking, object localization, environment mapping and human action/ activity recognition from both traditional and depth cameras located into the environment (Seo et al., 2015) or by considering first person point of view (Matsuo et al., 2014; Pirsiavash and Ramanan, 2012). Fall detection is one of the most faced health care issues and several computer vision techniques aiming at extracting human postures can be found in literature (Feng et al., 2014). These approaches focus on the analysis of the dynamic features (appearance change, shape deformation and physical displacement) that vary drastically in camera views while a person falls onto the ground (Yun and Gu, 2016). Frameworks for assessing the quality of a movement from skeleton data have been introduced to support clinicians in healthcare, to monitor rehabilitation of patients or to identify musculoskeletal disorders (Ho et al., 2016; Tao et al., 2016) . Recently, there has been also increasing attention in automatic eye blink detection for monitoring health (e.g., dry eye syndrome), for detecting face liveness or signs of sleepiness and for understanding some attempt of interaction from disabled people. In particular, recent approaches based on motion vectors analysis achieved

high invariance to appearance clutter such as glasses, head pose or facial expression (Fogelton and Benesova, 2016). The last frontier in this application context is the design and implementation of multisensory platforms (e.g., in the form of a smart mirror) integrating different sensors, including 3D optical scanner, multispectral cameras and gas detection sensor, collecting multidimensional data of individuals standing in front of it (Andreu et al., 2016). The goal is to enable users to self-monitor their well-being status over time and improve their life-style via tailored user guidance. A quite similar goal can be also achieved using mobile devices: for example in Marcon et al. (2016) the authors proved as mobile devices can fruitfully perform as an oral hygiene supervisor for kids. Food recognition is another emerging topic in health-oriented systems where it is used as a support for food diary applications. The goal is to improve current food diaries, where the users have to manually insert their daily food intake, with an automatic recognition of the food type, quantity and consequent calories intake estimation. In addition to the classical recognition challenges, the food recognition problem is characterized by the absence of a rigid structure of the food and by large intra-class variations (Farinella et al., 2016; Martinel et al., 2016). Others common assistive technologies dealing with the improvement of the comfort of person in daily life have been introduced through natural, creative and intuitive methods for communicating and by enabling remote control of devices (Rautaray and Agrawal, 2015). In this application context the exploitation of computer vision approaches allows the assistive systems to handle complex activities and to increase the possibility for the person to move freely and with a lower number of other wearable sensors. First Person Vision can be very useful in the context of daily activity monitoring. Wearable systems able to suggest to the user how to use an object or that are able to perform monitoring, summarization and temporal segmentation of daily actions and contexts in video (Damen et al., 2014; Furnari et al., 2015; 2016; Gurrin et al., 2014; Lee et al., 2012; Lee and Grauman, 2015; Lu and Grauman, 2013; Ortis et al., 2016; Poleg et al., 2014) can greatly help to support users in their daily life. Moreover, computer vision based assistive technologies can be adapted to reflect changes into the environment or to include new functionalities and, last but not least, they can satisfy anticipatory requirements with good accuracy (Forkan et al., 2015; Memon et al., 2014).

Autonomous robots are also making their way into common environments such as houses and offices to support daily living activities of humans. In context the possibility to make use of real time Human Robot Interaction (HRI) systems (to allow user communication with the robot in an easy, natural and intuitive gesture-based fashion) plays a central role and for this reason it is being deeply investigated (Canal et al., 2016). As an example of advanced research outcome, a domestic assistive robotic system has been proposed in Mollaret et al. (2016). Its main feature is that it only starts interaction with a user when it detects the userÆs intention to do so (i.e., it is not intrusive). This is achieved by multi-modal perceptions which include user detection based on RGB-D data, userÆs intention-for-interaction detection with RGB-D and audio data, and communication via user distance mediated by speech recognition.

3. Computer Vision techniques involved in AT tasks

In the previous section the most relevant technologies involving computer vision tasks have been introduced for each area of user's needs. In this section, for each of the aforementioned computer vision tasks, a literature review of the leading methods and techniques which have been already employed to address assistive issues is given.

3.1. Visual self-localization and mapping

Self-localization is the computational problem of keeping track of an agent's location within a known environment. Localization is actually answering the question "Where am I?". If the agent moves in an unknown (or dynamic) environment, while locating itself, it has to simultaneously construct and update a sparse or dense 3D model of the scene while traveling through it. In this case the problem becomes more challenging and it is referred with the acronym SLAM, standing for Simultaneous Localization And Mapping, that is referred as visual SLAM when it is based on visual inputs. Localization in a known environment is a well studied field in robotics and several approaches have been proposed in the past to address assistive issues. It is generally performed by using panoramic (commonly ceiling mounted) or perspective cameras opportunely positioned in the environment. In particular ceiling cameras are simple and cheap to implement, are not occluded and provide a large view of the environment in one image. Ceiling cameras for SLAM purposes are commonly used in Socially Assistive Robot since the on board camera resolution is often too low, as for example for the NAO[1] platform, one of the worldwide most widely used robot for assistive purposes. Solving the problem of localization means to identify global coordinates in the 2D plane and the orientation of the robot (i.e. the robot pose). The problem is often addressed by using probabilistic approaches that localize the moving agent with respect to a given map. Techniques such as extended Kalman filters (EKF), histogram filters or particle filters, often referred to as Monte-Carlo localization (MCL) (Rowekamper et al., 2012) are commonly exploited. The underlying processing pipelines of the different SLAM systems share the same scheme. The camera motion is tracked using frame-to-frame matching or feature-based approaches and the pose updating is determined by computing a relative transformation between a (partially) reconstructed world (i.e. the map) and a set of features or 3D point clouds obtained from the live camera. Frame-to-frame matching is, in general, inherently prone to drift whereas feature based tracking discards substantial parts of the acquired image data during the process of selecting relevant keypoints and, hence, is straightforward that the detectors and descriptors used in this process impact the accuracy of SLAM (Schmidt and Kraft, 2015). The descriptors usually used for SLAM issues are Scale-invariant feature transform (SIFT) and Speeded Up Robust Features (SURF), even if other solutions are under investigation by the community (e.g. Space-time Gabor, 3D histogram of oriented gradients) and the final choice is still being debated (Rivera-Rubio et al., 2015). Features are extracted from the incoming images and then they are matched against features from the previous images. Then a set of point-wise correspondences between any two frames is obtained and, based on these correspondences, the relative transformation between the frames is estimated using methods such as Random Sample Consensus (RANSAC) (Engelhard et al., 2011). In order to speed-up the matching process, a Kalman Filter that combines the matching results of current observation and the estimation of robot states based on its kinematic model can be exploited (Nguyen et al., 2015). A recent approach considered the place recognition step as a classification problem and proposed an efficient search space reduction considering only navigable areas where the user/robot can be localized (Sánchez et al., 2016). Sometime fiducial markers are used since they can be easily detected with low cost cameras (Coughlan and Manduchi, 2009) even from wearable or hand-held cameras along crowdsourced paths (Rivera-Rubio et al., 2016). Exploiting advances in imaging techniques, it has become quite practical to capture RGB sequences as well as

[1] Aldebaran Robotics, https://www.aldebaran.com.

depth maps in real time making use of devices such as the RGB-D cameras of Microsoft Kinect[2] and ASUS Xtion Pro Live[3] which provide additional information of object shape and distance compared to traditional RGB cameras. These technological advancements have led many researchers to investigate the possibility to create a novel approach to SLAM that combines the scale information of 3D depth sensing with the strengths of visual features to create dense 3D environment representations (Endres et al., 2012). RGB-D sensors are also used to detect "visual noun" features: signage, visual text, and visual icons that are proposed as a low cost method for augmenting environments or for navigation assistance (Molina et al., 2013). On the other side, the use of RGB-D cameras increases the computational load and then more computationally efficient algorithms have to be introduced in the SLAM pipeline (Lee and Medioni, 2015). The most popular algorithms for visual SLAM are present in the Robot Operating System (ROS),[4] an open source set of software libraries and tools for building robot applications.

3.2. Detection and tracking of objects

Object recognition for assistive technologies aims can be broadly divided into two categories: 1) marker-based approaches, requiring visual tags placed on the objects to be identified 2) marker-less approaches which utilize more complex sequences of algorithms to characterize information such as the objectÆs shape, texture, color, and other physical features (e.g., size) of the object to be identified (Jafri et al., 2014). Marker-based approaches perform recognition by capturing and analyzing an image of the marker which is usually designed to be simple to distinguish from other objects in the scene (e.g., a square of a specific color). Marker-based approaches require low computational power and storage space and they are ideal for tasks dealing with very cluttered scenes (e.g., identification of objects in room full of toys during experiments in the context of autism disorder). However, marker-based systems require careful preliminary selection of markers and the correct placement of them on the objects. Moreover visual markers have to be in line-of-sight of the camera, otherwise, they will not be detected. Furthermore, the visibility of these markers may be unappealing from an aesthetic perspective. These limitations can be overcome using marker-less approaches. From a more technical point of view the algorithms dealing with marker-less object detection and tracking can be roughly divided into two subcategories. In the first one, tracking is performed by detecting an object frame by frame (traking by detection). In the second subcategory, after the detection (based on a priori model representation of the objects) the tracking is performed usually on the basis of an object model that is dynamically updated during the tracking. Usually for the detection step, a variety of visual features can be employed to represent the object under consideration: Scale-invariant feature transform (SIFT), Speeded Up Robust Features (SURF), Edge/Corner, Gabor kernels, Wavelet Transform, random ferns, Binary Robust Independent Elementary Features (BRIEF), Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), Local Ternary Patterns (LTP), Census Transform Histograms (CENTRIST), Bag-of-Visual-Words (BOVW) (Wang et al., 2014b; Yu et al., 2015). The design of the above local image features used for object detection make them robust to geometric and photometric variations, such as rotation, scaling, viewpoint and illumination change. Sometimes multi-resolution and multi-orientation approaches are used to improve detection accuracy (Matusiak et al., 2014) but, unfortunately, these strategies are accompanied by increased computational costs. People are one of the most challenging classes for object to be detected, mainly due to large variations caused by the intrinsic deformability of the shape and high variability in appearance. For this reason, a combination of multiple features as well as motion features can be used to assure the detection performance required in an assistive context (Diraco et al., 2015).

Independently from the involved representation strategies, the model of an object is used to perform matching with respect to a dataset of examples, or it is given as input to a machine learning technique (e.g., artificial neural network, Support Vector Machine, boosting, probabilistic approaches, decision trees, random forests...) after a training stage (Wei et al., 2014).

Semantic segmentation is another common way to address this problem. In the first stage, every pixel is classified based on its appearance. The output of the first classifier in a specific region around every pixel is summarized by a voting histogram feature and given as input to a second classifier (Haltakov et al., 2016; Raví et al., 2016). Interactive object learning methods using scanty human supervision are alternative ways to handle unknown and difficult recognition cases (Villamizar et al., 2016). Multimodal applications which combine RFID with visual cues of the objects have been also proposed with the aim of enhancing the ability of the object detection and tracking systems (Jafri and Ali, 2014). Once an object has been identified in the scene it is usually useful to track it over consecutive frames by considering a set of keypoints as well as its appearance. Keypoint based (such as Kalman (Takizawa et al., 2015) or particle filtering (Yan et al., 2011)), Kernel-based (Thakoor et al., 2015) and shape-based (Sivalingam et al., 2012) tracking solutions have been exploited in the assistive context. Finally tracking in egocentric videos has been faced in Aghaei et al. (2016) by introducing several features that help in increasing robustness even in photo-streams acquired by cameras with lower frame rates and narrower fields of view. It must be taken into account that object tracking in egocentric photo-streams is a different problem from the tracking in conventional videos in several aspects. Conventional tracking facilitates itself with the assumption of temporal coherence, while temporal coherence does not hold for egocentric photo-streams. Moreover, in egocentric photo-streams, the appearance of the target as well as its position may change drastically from frame to frame. In addition, due to changes in the camera field of view caused by body movement of the camera wearer, background modeling becomes a more challenging issue.

3.3. Human activity recognition

Human activity recognition from a sequence of images, either captured by RGB cameras or RGB-D sensors, is one of the most challenging topic in computer vision (Tsoukalas and Bourbakis, 2013). In the most common taxonomy, human activities are composed by a sequence of primitive displacements named actions (if they involve the movements of the full body as in the case of walking, sitting, jumping, bedding, etc.) or gestures (if they involve the displacement of only one or more parts of the human body as for example pointing at objects, shaking head, clapping hands, head nodding, etc.). Many of the computer vision approaches addressing this problem, employ descriptors extracted from preliminary detected human silhouettes or human body parts (e.g., the upper body limbs and the hands) (Delibasis et al., 2014). The detection and tracking of the human silhouettes and human body parts follows the general operative framework adopted for object detection and tracking (see Section 3.2). Recent trend in assistive human activity recognition research is the exploitation of low cost RGB-D cameras (e.g., Microsoft Kinect) which are suitable for generating skeleton model of a human with 15 body joints

[2] www.microsoft.com/en-us/kinectforwindows/.
[3] www.asus.com/Multimedia/.
[4] www.ros.org.

positions and their relative orientations. Considering the skeleton features the human activities can be recognized (Piyathilaka and Kodagoda, 2015). In general, once the human silhouettes or human body parts have been detected, the recognition of human activities is then performed in three different steps: segmentation, representation and classification. In general, in the systems supporting human needs, the segmentation of human activities is performed using temporal sliding windows that can have fixed or dynamically derived lengths (Okeyo et al., 2014). The variations in the body configuration can be then represented with spatio-temporal features which focus on the representation of the shape and motion as a function of time, i.e. creating space-time volumes. The recognition of human activity is finally achieved by comparing the observed space-time volumes with predefined ones that can be learned by training data or built using domain knowledge (Giakoumis et al., 2014). Common machine learning techniques used to recognize models of human activities are Support Vector Machine, K-nearest neighbor, Naive Bayes, Hidden Markov Model, Conditional Random Fields, Artificial Neural Network and Decision Tree (DT) (Tong et al., 2015). In medical application context, statistical approaches have been used to encodes the motion and position statistics of densely extracted trajectories from a video in order to recognize actions (Iscen et al., 2015). In some studies, classifiers are combined in different ways, thereby creating multi-layer or hierarchical classification (Shoaib et al., 2015), whereas other ones proposed relational graph of visual words for activity analysis for smart home security (Zhang et al., 2015b). For gesture based interfaces, the precision (high true positives and low false positive rates) has to be assured while keeping the natural feeling of interpersonal communication. This can be achieved using graph matching where the nodes in the graph are assigned at different hierarchy levels, relative to their importance in the matching process (Li and Wachs, 2014).

Assistive systems must assure reliability of the adopted activity recognition models and therefore the model should also be capable of detecting and avoiding false assignments. To this aim approaches able to measure the reliability of the assigned action label by using a confidence score to highlight activities with high or low confidence have been recently proposed and applied into a smart home assistive domain Fahad et al. (2015).

Advances in wearable technologies are facilitating the understanding of human activities using first-person vision (FPV) for a wide range of assistive applications. For example in Abebe et al. (2016) a robust motion-feature (RMF) that combines grid optical flow-based features (GOFF) and video-based inertial features (VIF) has been tested using support vector machine as classifier. Recognizing finger writing in mid-air is another useful input tool for wearable egocentric camera. Unfortunately it suffers from the problem of fast viewpoint/scale changes and then, to achieve this challenging task, contour-based view independent hand posture descriptor that serves both posture recognition and fingertip detection was proposed. As to recognizing characters from trajectories, a Spatio-Temporal Hough Forest that takes sequential data as input and perform regression on both spatial and temporal domain was used (Chang et al., 2016).

An even more challenging problem in this context is the early recognition of activities. Early recognition, which is also known as activity prediction, is an ability to infer an ongoing activity at its early stage and it is crucial, especially in the AT research field, since fast reactions can be activated. The activity prediction approaches model how feature distributions of activities change as observations increase. Integral histogram representations of the activities are constructed from training videos and then recognition methodologies extending the prediction algorithm to consider the sequential structure formed by video features are used (Ryoo et al., 2015).

3.4. Biometrics

Biometrics techniques can be largely divided into traditional (hard) and soft biometrics. Hard biometrics is aimed in recognizing unique and permanent personal characteristics of a person and, in the AT context, can be exploited to create personalized setups, restrict delegation access rights as well as to discourage fraudulent access or impersonation of users. Moreover, biometric authentication technologies facilitate the remote access to electronic health records for both patients and other stakeholders, and provide a secure method for encryption of personal data. The traditional authentication modalities involving computer vision are based on hand, facial or ocular image recognition (Unar et al., 2014). Unfortunately, traditional biometrics necessitate of considerable user cooperation and the acquisition of the characteristic is intrusive. This is a drawback especially in the AT application field since patients are often non collaborative and get scared when interacting with the authentication systems.

Soft biometrics are human characteristics providing categorical information about people such as gender, ethnicity, height, weight, skin color, eye color, hair color and SMT (scars, marks and tattoos). In contrast to hard biometrics, soft biometrics provide some vague physical information which is not necessarily permanent or distinctive. Such soft biometric traits are usually easier to capture, do not usually require cooperation from the subjects and for this are perfectly suitable for assistive contexts. Research on soft biometrics related to assistive technologies is still in its infancy despite the vast potential applications that range from the support to traditional authentication schemes to the design of agents able to successfully engage and interact with humans in an adaptive way (Zhang et al., 2015a). Soft biometrics modules working on RGB cameras concentrate on the human face since it contains a large amount of useful information for the scope. The most common procedural scheme consists in a preliminary step that detect human faces by a Boosted Cascade of Simple Features (e.g. projection on Haar Basis functions) eventually evaluated along multiple frames. The face preprocessing component is implemented by cropping the facial areas from images. The detected faces may be then downsampled to boost execution time and then the classification of soft-characteristics is performed. In this research area the most used classifiers are k-Nearest Neighbor (kNN), Support Vector Machines (SVMs), Bayesian Networks and decision tree which work taking as input either information extracted for instance through Principal Component Analysis (PCA), Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), local ternary patterns (LTP), active appearance models (AAM) or geometrical information extracted by automatic facial feature point localization (Gabor-Facial Point Detector, Active Shape Models, etc.) (Bosch et al., 2015; Carcagnì et al., 2015b; Salah et al., 2010). In alternative to facial traits, also 3D body metrics can be extracted from a number of networked RGB-D devices. Skeleton joints have been used in literature to compute height, shoulder breadth, hip breadth, head length and the proportion ratios of person and also used for gender and age estimation (Sandygulova et al., 2014).

3.5. Head pose/gaze estimation

The automatic recognition of the orientation of the head, eventually strengthened by the more accurate estimation of the gaze direction, is crucial in a number of AT applications as human-computer interaction, autism diagnosis, monitoring of social development, depression detection and human behavior analysis. The most common used paradigms for head pose estimation are called Generalized Adaptive Viewbased Appearance Model (GAVAM) and Constrained Local Models (CLM). The Generalized Adaptive View-based Appearance Model has three main advantages: 1) the au-

tomatic initialization and stability of static head pose estimation, 2) the relative precision and user-independence of the differential registration, and 3) the robustness and bounded drift of keyframe tracking (Mead et al., 2013). The Constrained Local Models follows the Active Appearance Model approach trying to model faces via a statistical description based on a set of landmarks. Shapes of faces are deformed iteratively according to the landmarks positions in order to find a best fit with the actual face image. Both GAVAM and CLM can be used with depth data (Papoutsakis et al., 2013), despite the use of RGB images is often preferred. In the context of assistive technologies, mobile platforms have been also used as acquisition devices (Anzalone et al., 2014). Starting from the information about the head pose, the Focus Of Attention, FOA (i.e. what or whom a person is looking at, sometimes referred as Point of Regard or Point of Gaze) can be also estimated. FOA is important for robots or conversational agents interacting with multiple people, since it plays a key role in turn-taking, engagement or intention monitoring. In addition to the aforementioned applications, an eye tracking system can be applied to help people with severe disability to manipulate computers or to pursuit ocular movements in the diagnosis or evaluation of diseases (e.g., multiple sclerosis). Approximating FOA from head pose creates ambiguities since the same head pose can be used to look at different FOA targets and this can become a limitation to reach a predefined assistive goal. This limitation can be overcome using information about the localization of the eyes and gaze estimation, enabling a more accurate FOA estimation (Sheikhi and Odobez, 2015). In terms of the physical structure, systems enabling the localization of the eye are grouped into head mounted systems or remote systems. Other categorizations could be considered taking into account wearable or non-wearable modality, as well as infrared-based or appearance-based information. Systems that operate remotely are more suited for assistive applications since technology providers cannot always rely on the user's cooperation. Most of the systems that work remotely use an artificial infrared (IR) light source that produces a corneal reflection into the center of the pupil. IR based systems are very accurate, tolerant to head pose variation, and computationally efficient but they have also drawbacks. Indeed, the harm of IR for human eyes have to be considered, they have to be calibrated and/or have specific requirements for the imaging device (Topal et al., 2014). Appearance based systems operate, instead, on the images acquired by RGB or RGB-D cameras without any additional hardware. They generally operate on images acquired from static cameras. The implementation on mobile devices of these systems is a challenging task (Amudha et al., 2015). The basic idea is to find the face within the image and then locate eye centers using geometric information (symmetry or curvature of the iris), the analysis of image gradients, the curvature of isophotes or a combination of them (Leo et al., 2014). Finally combining estimates of Head Pose and Eye Gaze, it is possible to get the Visual Focus of Attention for HCI purposes, without any special needs in terms of hardware, or knowledge of internal camera or set-up parameters (Asteriadis et al., 2014). In the light of this, head pose estimation is becoming a key point for successful gaze estimation (Ariz et al., 2016) alleviating any assumptions of stationary head movement without requiring prolonged user co-operation prior to gaze estimation (Cristina and Camilleri, 2016).

After the location of the eye centers, the gaze direction is generally estimated by either interpolation or 3D models. Interpolation methods use general purpose equations, such as linear or quadratic polynomials, to map the image data to gaze coordinates. 3D model-based techniques, on the other hand, directly compute the gaze direction based on a geometric model of the eye (Cazzato et al., 2014).

4. Critical assessment of AT: users, medical, economical and social perspectives

As discussed in the previous sections, the great advances in computer vision allow to obtain new and more powerful assistive devices and to improve existing ones, thereby increasing the potential benefits that can accrue. However, at the same time, this fruitful evolving brings out the need to take proactive and sophisticated forms of assessment of AT. Critical assessment of AT is a not trivial process that can take a long time considering that there are several actors involved and that it is part of the iterative and incremental development process. The central components in this process are the end-users: the assessment of satisfaction of the end users is a critical step regardless the underlying technology, but it is particularly challenging in the case of devices involving computer vision tasks since the users benefit of high level information that can require intermediate levels of interpretation and transduction which has to be assessed in their turn (for example a device to help mobility of blind persons can require a communication device based on audio. It is not straightforward to separate the assessment of the functioning from that of the communication modalities). It follows that computer vision based assistive devices generally require complex assessment processes and there is a danger that they become less accessible, with one of the greatest problems being that individuals stop using the devices (it was estimated that about one-third of all assistive devices are abandoned). The process of assessment itself is costly, given the number of professionals who potentially need to be engaged, and this issue has to be taken into account while defining the device development process (Federici and Scherer, 2012).

Recently, to investigate perceptions and self-perceptions of AT users, daily studies of groups of participants have being conducting involving both people with disabilities and people without disabilities. The goal is to explore the types of interactions and perceptions that arise around AT use in social and public spaces. The preliminary results pointed out that AT form or function influence social interactions by impacting self-efficacy and self-confidence. When the design of form or function is poor, or when inequality between technological accessibility exists, social inclusion is negatively affected, as are perceptions of ability. These studies contribute a definition for the ôsocial accessibilityö of AT and subsequently offer Design for Social Accessibility (DSA) as a holistic design stance focused on balancing an AT user's sociotechnical identity with functional requirements (Shinohara and Wobbrock, 2016).

In general, good design, usability and accessibility factors are undoubtedly needed for assistive technology to properly perform for end-users. Unfortunately these key factors are not sufficient. Indeed, the user's lifestyle and aspirations have to be taken under consideration to get and consolidate a positive user's evaluation. This means that even a great designed technology can be poorly evaluated. This consideration has pushed the definition of a number of complex frameworks for the evaluation of the assistive technology outcomes.

One of the commonly used framework to model the outcomes of an assistive technology is the Matching Persons and Technology (MPT) Model (Fuhrer et al., 2003). The evaluation procedure requires both the provider and user to complete slightly different versions of forms, followed by a discussion of the results and action. This method draws on the medical model of disability. Starting from "limitations" on functioning, it identifies goals and technological modules that could be exploited by taking also into account characteristics of the person (experiences and attitudes to technologies, degree of satisfaction with different aspects of life) and environment. This way the model tries to predict behavioral tendencies that could lead to inappropriate use or abandonment of the recommended technologies.

Another relevant evaluation tool, proposed by the Consortium for Assistive Technology Outcomes Research (CATOR), considers a taxonomy of assistive products reflecting combinations of user population, AT type, services, and context for use. It builds on the International Classification of Functioning Disability and Health (ICF) (WHO 2001) and is based on three descriptors: effectiveness, social significance and subjective well-being (Jutai et al., 2005).

One more example is the SETT Framework aimed to promote collaborative decision-making for implementation and evaluation of effectiveness of AT for students (SETT is an acronym for Student, Environments, Tasks, and Tools) (Zabala, 1995).

As previously mentioned, the service provider and the end-user are not the only actors in the process of assessment of AT. Medical perspective is another key factor. Medical personnel, including both doctors and care personnel, are more reluctant and address different aspects which need to be taken into account. This distrust of assistive technologies does not differ in the specified age and gender groups (Wilkowska et al., 2010) and it can be, for example, based on the higher probability of false alarms (in case of remote monitoring) or in the feeling that a machine (e.g. a robot) can make at their place what they are skilled to do. Additional doctors barriers are low usability of technical devices, assumed difficulties in handling the devices, and low technical competence which might be the reason for their view that AT are more time-consuming in contrast to the conventional approaches (Ziefle et al., 2013).

However, these drawbacks are almost overcome when using computer vision tools since, on the one side, the care personnel can directly check the actual situation through the image streams (false alarm can be then easily filtered without expensive efforts) and, on the other side, their use (e.g., in remote surgery or telemedicine) can be guided by graphical or interactive tools implementing user-friendly interfaces. As an example of recent investigations about the different AT assessment of end-users and medics is reported in Oliver et al. (2015). The authors proposed the use of RGB-D sensors in motor rehabilitation of patients suffering acquired brain injury and a qualitative evaluation of how the stakeholders of the proposed AT environment (injured patients and therapists) use and perceive the ambient assisted rehabilitation room was carried out. Three factors are then extracted and results demonstrated that acquired brain injury patients are more concerned with the aesthetic and operability factor than therapists. On the other side, therapists demand the protection aspect and rehabilitation room more than others do. From this example emerged that the medical perspective has to be also included in the assessment process in order to achieve a careful and far-sighted planning of AT.

Another aspect that plays an important role is related to the economic evaluation of AT. In fact, even the most developed welfare state systems allocate a limited amount of economical resources to provide AT. This implies that social services, health care and other organizations supplying assistive technology to individuals must take operational decisions on the basis of accurate evaluations to ensure that all costs and benefits are included. However, the economical evaluation is not a straightforward process since it involves both practical and ethical problems associated with some of the possible costs and benefits, such as improvements of the quality of life and new or foregone opportunities. Sometimes costs are challenging to determine since can have high levels of uncertainty (costs may be hidden or not apparent until later) which make practically impossible an immediate and concrete quantification. To bypass this bottleneck, it has been suggested to use monetary valuations of costs and non-monetary evaluation of benefits. This is, for example, the key idea behind the CERTAIN study that proposed a model of cost-outcome analysis for assistive technology (Andrich et al., 1998). A practical way to make the economical assessment task easier is to focus cost-effectiveness studies on group of people with specific needs. This has been largely done for AT supporting people with dementia (Bowes et al., 2013) but it is absent for other groups of people.

Finally, also social aspects of AT must be taken under consideration since there may be concerns: 1) about how the data acquisition devices may affect the privacy or freedom of a person, in particular in the case of computer technologies that rely on sharing visual information. 2) about the impact on society, for example, the fear that it may be used to cut back services and reduce human contact or that some devices may be used to do things a person is still able to do for themselves which may make their problems worse or that they may make things more complicated or beyond the abilities of the person or that they may help foster a one-sided focus on a person's problems and not on their existing strengths. In particular, the ethical and social implications of using computer vision are critical and require more in-depth study. It has become clear that the issue of acceptance of AT that rely on computer vision is becoming more and more important and that it deserves a deeper discussion. More in-depth studies need to be conducted which explore stakeholders critical issues such as: determining the types of home monitoring technology that users would accept into their homes; in which daily life tasks such technology is acceptable; who will view data collected from such systems and during which type of circumstances.

Furthermore, the ethics involved in developing and using such systems with potentially vulnerable populations needs to be studied. For example, who makes the decisions for this population on when and how to use such technology? Questions related to the future of this technologies also need to be addressed, such as how will these ethical issues change in the future (Mihailidis et al., 2004).

Summing up, more accurate studies are still needed to acquire statistically significant data and then to get systematic approaches to quantify the key assessment factors of AT (acceptance level , cost and ethical implications) (Lenker et al., 2013). Anyway it is established that, in the development of an assistive technology product, the user centered design practices take an important role (Pawluk et al., 2015a). The opinion of the users for the determination of the real needs during the design process through focus group and the assessment via preference surveys after an experimental session are useful to make the final assistive product more effective (Devyver et al., 2011b). Participatory Action Design techniques (Cooper, 2007) which involve stakeolders in the design process (i.e., users, care providers, industry, etc.) from storyboarding to final real-world testing should be adopted to reach the assistive technology goals.

5. Open challenges

By analyzing the state of the art reported in previous section some still open technological challenges emerge. These challenges are explained in this section and, for each of them, some methods addressing the underlying problems which have been recently presented in literature are briefly discussed, explaining also how they could allow a real improvement of the assistive outcomes in the short and medium term. This discussion about present challenges in AT and possible ways to address them needs to start from a more general consideration about what is happening in the artificial intelligence community where deep learning (DL) is making major advances in solving problems that have resisted the best attempts for many years. It has turned out to be very powerful to discover structures in high-dimensional data and it is therefore applicable to many domains of AT. In addition to beating records in image and speech recognition, it has produced extremely promising results for various tasks in natural language

understanding, particularly topic classification, sentiment analysis, question answering and language translation (LeCun et al., 2015). Its very likely that assistive technologies will be soon the beneficiary of this research trend. In particular deep-learning-powered computer vision has the potential to become the next major step in vision correction. Unfortunately, as reported later in this section, this desired road from laboratory tests to exploitation in real contexts is rugged since research on deep architectures is still young and many questions remain unanswered. A few examples among many: DL takes advantage of large datasets which are not easy to build (methods based on pre-trained nets and unlabeled data are yet in their infancy) and, moreover, imperfections in the training phase make them vulnerable to adversarial samples. Another issue concerns the definition of the network model and the limited understanding of the optimization process. For these and many other reasons, most of the challenges involved in AT need further investigation that has to go beyond DL applications even if, as reported in the following, DL paradigm has sometimes already been proposed as an appealing solution. In what follows, some of the weaknesses of the most used approaches involved in AT are highlighted and for each of them the most recent and attracting Computer Vision solutions on the horizon are pointed out.

Going into details, concerning the self-localization problem, Monte-Carlo based approaches suffer in achieving long-term localization since it involves several additional issues, like handling of variant environments, error recovery, efficient place recognition, possible sensor occlusions, repeated occurrences of similar features due to repeating structure in the world (e.g., doorways, chairs, etc.) and missing associations between observations. Moreover for large maps, the performance of pure image matching techniques decays in terms of robustness and computational cost. This shortcomings may be addressed using recent introduced hierarchical data association methods that keep multiple associations per particle. They estimate a candidate pose using few correspondences between features of the current camera frame and the feature map. The initial set of correspondences is established by proximity in feature space. The initial estimated pose is then used in a second step to guide spatial matching of features in 3D space, i.e., searching for associations where the image features are expected to be found in the map. A RANSAC algorithm is used to compute a fine estimation of the pose from the correspondences (Gamallo et al., 2015). Another possible way to increase localization performances could be the extension of Monte Carlo Localization method to 3D environments using inexpensive RGB-D cameras (Fallon et al., 2012). The use of a Convolutional Neural Network (CNN) can bring functional improvement to detect specific objects and recognize their orientation (Poggi et al., 2015) or to regress the cameraÆs orientation and position (Kendall et al., 2015).

When performing SLAM, frame-to-frame matching approaches have experienced a large amount of drift that can be partially explained by the fact that the visible scene is always very local and that only a rough intrinsic calibration was available. On the other hand SLAM approaches relying only on feature matching are less accurate due to their inherent sparseness. To overcome these limitations the dense tracking and mapping methods (Kerl et al., 2013), the combination with the metrical information given by the robot odometry (Bazeille et al., 2015) or the improvement of the estimation by means of the detection of moving objects thanks to a dense scene flow (Alcantarilla et al., 2012) can be adopted, also with limited computational resources (Panteleris and Argyros, 2014). A recent attempt investigated if the awareness of ego-motion (i.e. self motion) can be used as a supervisory signal for feature learning on the tasks of scene recognition, object recognition, visual odometry and keypoint matching (Agrawal et al., 2015). Although it doesn't work very well yet, one can imagine the effects of hooking this new strategy (perhaps trained with a lot more data?) into SLAM pipeline to obtain much more robust associations between neighboring frames or a frame with its nearest keyframe, than edges alone can provide.

From the analysis of the approaches performing object detection in AT domain it emerges that they find difficulties to handle the variations of the objects due to scale, rotation and deformation (especially when a person has to be detected). Active learning has recently emerged as a powerful tool for building robust systems for object detection (Sivaraman and Trivedi, 2014) and can be used also in AT context. As an alternative multi-resolution and multi-orientation detectors can be exploited and the inherent computational costs can be circumvented by a reliable prediction of image structures across scales that can rely on the assumption of the fractal structure of much of the visual world (Dollar et al., 2014) or on the data-driven estimation of the likelihood that a region being an object (Kang et al., 2015). Also high level image features can be used to represent complex real-world image by collecting the responses of many object detectors at different spatial locations (Li et al., 2014). Some recent papers propose also to use a large library of textured 3D object models (publicly available on the Internet) to implicitly represent both the 3D shape of the object class, as well as its view-dependent 2D appearance (Aubry et al., 2014). Semi-supervised learning, in which the training set consists of a large number of unlabeled examples and a small number of labeled ones, are also gaining attention to this aim (Sun and Savarese, 2014). Some authors propose the use of local and global shape descriptors in order to improve the accuracy in shape classification and retrieval (Battiato et al., 2015). Object classification and object detection have rapidly progressed with advancements in CNN and the advent of large visual recognition datasets. Modern object detectors predominantly follow the paradigm in which first an object proposal algorithm generates candidate regions that may contain objects, second, a CNN classifies each proposal region. Most recent detectors follow this paradigm and they have achieved rapid and impressive improvements in detection performance (Zagoruyko et al., 2016).

Concerning object tracking, significant and rapid appearance variation due to noise, occlusion, varying viewpoints, background clutter, illumination and scale changes pose challenges to this task. Non-rigid object tracking, in which the object model evolves during the tracking process as the appearance of the object changes, is one possible solution to achieve good performances. This can be achieved by adaptive appearance models (Babenko et al., 2011; Huang et al., 2015), kernel-based region covariance descriptor (Wu et al., 2015) or by exploiting contextual information (Di Lascio et al., 2013). Extending current deep learning architectures to the tracking problem is not a straightforward process instead. Anyway, a recent work has pursued attempts to use them for tracking, showing some promising results (Zhai et al., 2016). In particular, an online object tracking algorithm based on deep neural networks, in which the network learns the probability densities of appearance of the target and its surroundings and updates itself adaptively to new observations was presented.

Moving on to the task of recognition of human activities, the major problems are occlusions, shadows and background extraction, lighting condition variations, viewpoint changes and, last but not least, the possibility that the same activities is executed differently from different persons or even by the same person. To make the recognition more robust, the analysis of human gesture/action and behaviors has been recently faced considering a new perspective, that brings in notions and principles from the social, affective, and psychological literature, and that is referred as Social Signal Processing (SSP). SSP integrates Computer Vision and Pattern Recognition methodologies with nonverbal cues, like face expressions and gazing, vocal characteristics, relative distances in the space and so on (Cristani and Murino, 2014).

Another possible way forward leads to vision-based systems combining RGB and depth descriptors: they have been already successfully used to classify hand gestures for a human-machine interface application in the car (Ohn-Bar and Trivedi, 2014) or to describe the articulated human body that has a large number of kinematic joints (Wang et al., 2014a). Another relevant improvement in this research field has been the proposition of the sparse coding-based temporal pyramid matching approach (ScTPM) for feature representation. Due to the pyramid structure and sparse representation of extracted features, temporal information is well kept and approximation error is reduced. In addition, a novel Center-Symmetric Motion Local Ternary Pattern (CS-Mltp) descriptor has been proposed to capture spatial-temporal features from RGB videos at low computational cost (Luo et al., 2014). Additional researches focus on the evidence that activities are normally related to the concept of interaction between a person with one or more people or between one or more people with objects of the surrounding environment (Chaquet et al., 2013). Several recent approaches, on the other hand, are based on learning personûobject interactions and saliency maps in images, investigating the possibility and applicability of identifying action-specific points or regions of interest in still images based on information extracted from video data (Eweiwi et al., 2015). Another possibility is to select a subset of discriminative frames from a video to improve the performances of detection and recognition of human interactions (Sefidgar et al., 2015). Finally, early prediction of ongoing human activity has become more valuable in a large variety of time-critical assistive applications (especially related to heath monitoring). To build an effective representation for prediction, human activities can be characterized by a complex temporal composition of constituent simple actions and interacting objects. Frameworks for long-duration complex activity prediction by discovering three key aspects of activity (Causality, Context-cue, and Predictability) have been also proposed (Li and Fu, 2014). Deep learning has the potential to have significant impact on human activity recognition. Gkioxari and Malik (2015) extended this approach for action localization. Donahue et al. (2015) proposed an end-to-end trainable recurrent convolutional network which processes video frames with a CNN, whose outputs are passed through a recurrent neural network. In this field the state of the art is represented by the approach proposed in Rahmani et al. (2016) where a deep fully-connected neural network that transfers knowledge of human actions from any unknown view to a shared high-level virtual view by finding a non-linear virtual path that connects the views is proposed.

Detecting predefined facial feature points in a human face is a well studied problem but, unfortunately, it is still an open challenge under unconstrained environments, with variations of illumination, expression, head pose, as well as partial occlusions. Methods which extract biometric measures using depth sensors defining a representative model and fitting a 3D shape context descriptor within an iterative matching procedure (Madadi et al., 2015) can be used to improve recognition (Azazi et al., 2015). Alternatively facial feature points under challenging conditions can be located by enhancing the classification process, for example making use of probability outputs trained to provide the observation probability of each facial feature point. The problem can be then handled by maximizing the posterior which combines the prior and observation probability. This approach has been effectively adopted within the framework of Bayesian Inference using Support Vector Machines with observation probabilities and distribution of face shape, which serves as the prior, that are both approximated with Gaussian Mixture Models (Wang et al., 2015). For detecting facial landmarks that remain invariant across head poses a fast cascade regressors (Z-face) (Jeni et al., 2015) or a 3D Morphable Model (3DMM) combined with the Structure from Motion (SfM) method (Jo et al., 2015) can be also considered. Many recent works on face recognition have proposed numerous variants of CNN architectures for face recognition. However, large scale public datasets have been lacking and, largely due to this factor, most of the recent advances in the community remain restricted to Internet giants such as Facebook and Google (Parkhi et al., 2015).

Head/gaze pose estimation systems used for AT issues still have some shortcomings in computational efficiency and precision. To overcome them, on the one hand, approaches that combine demographic recognition (gender) and behavior analysis (gaze) have been proposed in order to create a user-centered HCI environment by better understanding the needs and intentions of its users (Zhang et al., 2016). On the other hand, more sophisticated algorithmic strategies have been investigated i.e. a new type of concise 9-dimensional local descriptors with Fisher vectors that have been introduced in literature (Ma et al., 2015) to describe the head images. Concerning the estimation of the gaze direction an accurate and efficient eye detection method using the discriminatory Haar features (DHFs) and a new efficient support vector machine (eSVM) have been recently published (Chen and Liu, 2015). Ensemble methods to deal with the problem have been also experimented (Markuš et al., 2014). Gaze estimation models exploiting RGB-D cameras which rely on geometric generative gaze estimation have been recently also discussed in literature (Mora and Odobez, 2014). Perhaps the most promising approach is to train a machine-learning algorithm called a Constrained Local Neural Field (CLNF) to recognize gaze direction by studying a large database of images of eyes (even artificially created) in which the gaze direction is already known. Another possible way to improve computer vision based assistive systems could be the exploitation of multimodal inputs to combine information obtained from visual data with auditory and physiological signals and more in general with other signals acquired from sensory different than a camera (e.g., gyroscopes, accelerometers, etc.). An example of problem where a combination of the different signals can help to improve reliability of the assistive system is fall detection (Mubashir et al., 2013). Attempts to exploit multimodal information are already present with the aim at preventing risks for patients, by using multiple sensors providing real-time information on their status and activity (DemCare, 2016).

6. Conclusions

This paper focused on the computer vision components exploited in the existing assistive technologies (AT). In particular, a “task oriented” way to categorize the works related to AT is proposed starting from the consideration that the same computer vision task (e.g. object detection) can be part of technologies dealing with quite different user's needs and thus, by grouping the assistive technology works dealing with the same CV task could help the reader to better identify the research lines to improvement of technologies. The categorization considers the list of human needs developed by the world Health Organization (WHO) and the each Computer Vision tasks has been associated with the related assistive technologies. Eight relevant computer vision tasks employed in AT have emerged (localization, SLAM, Object Detection, Object Tracking, Human Activity Recognition, Biometrics, Head Pose and Gaze Estimation) and the related approaches in the assistive technologies literature have been summarized. Finally the recent advancements of the considered computer vision tasks, but applied in others application domains, have been discussed to make aware the reader about the new solutions which can be used to better accomplish of the user needs in the assistive domain.

References

Abebe, G., Cavallaro, A., Parra, X., 2016. Robust multi-dimensional motion features for first-person vision activity recognition. Comput. Vision Image Understanding 149, 229–248.

Aghaei, M., Dimiccoli, M., Radeva, P., 2016. Multi-face tracking by extended bag-of–tracklets in egocentric photo-streams. Comput. Vision Image Understand. 149, 146–156.

Agrawal, P., Carreira, J., Malik, J., 2015. Learning to see by moving. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 37–45.

Alcantarilla, P.F., Yebes, J.J., Almazán, J., Bergasa, L.M., 2012. On combining visual SLAM and dense scene flow to increase the robustness of localization and mapping in dynamic environments. In: Robotics and Automation (ICRA), 2012 IEEE International Conference on. IEEE, pp. 1290–1297.

Amudha, J., Nandakumar, H., Madhura, S., Reddy, M., Kavitha, N., 2015. An android-based mobile eye gaze point estimation system for studying the visual perception in children with autism. In: Computational Intelligence in Data Mining - Volume 2: Proceedings of the International Conference on CIDM, 20–21 December 2014. Springer India, New Delhi, pp. 49–58.

Andreu, Y., Chiarugi, F., Colantonio, S., Giannakakis, G., Giorgi, D., Henriquez, P., Kazantzaki, E., Manousos, D., Marias, K., Matuszewski, B.J., Pascali, M.A., Pediaditis, M., Raccichini, G., Tsiknakis, M., 2016. Wize mirror - a smart, multisensory cardio-metabolic risk monitoring system. Comput. Vision Image Understand. 148, 3–22.

Andrich, R., Ferrario, M., Moi, M., 1998. A model of cost-outcome analysis for assistive technology. Disability Rehabil. 20 (1), 1–24.

Anzalone, S.M., Tilmont, E., Boucenna, S., Xavier, J., Jouen, A.-L., Bodeau, N., Maharatna, K., Chetouani, M., Cohen, D., Group, M.S., et al., 2014. How children with autism spectrum disorder behave and explore the 4-dimensional (spatial 3d+ time) environment during a joint attention induction task with a robot. Res. Autism Spectr. Disord. 8 (7), 814–826.

Ariz, M., Bengoechea, J.J., Villanueva, A., Cabeza, R., 2016. A novel 2d/3d database with automatic face annotation for head tracking and pose estimation. Comput. Vision Image Understand. 148, 201–210.

Asteriadis, S., Karpouzis, K., Kollias, S., 2014. Visual focus of attention in non-calibrated environments using gaze estimation. Int. J. Comput. Vision 107 (3), 293–316.

Aubry, M., Maturana, D., Efros, A.A., Russell, B.C., Sivic, J., 2014. Seeing 3d chairs: exemplar part-based 2d-3d alignment using a large dataset of cad models. In: Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. IEEE, pp. 3762–3769.

Avola, D., Spezialetti, M., Placidi, G., 2013. Design of an efficient framework for fast prototyping of customized human–computer interfaces and virtual environments for rehabilitation. Comput. Methods Programs Biomed. 110 (3), 490–502.

Azazi, A., Lutfi, S.L., Venkat, I., Fernández-Martínez, F., 2015. Towards a robust affect recognition: automatic facial expression recognition in 3d faces. Expert Syst. Appl. 42 (6), 3056–3066.

Babenko, B., Yang, M.-H., Belongie, S., 2011. Robust object tracking with online multiple instance learning. Pattern Anal. Mach. Intell. IEEE Trans. 33 (8), 1619–1632.

Battiato, S., Farinella, G., Giudice, O., Puglisi, G., 2015. Aligning shapes for symbol classification and retrieval. Multimedia Tools Appl. 1–19.

Bazeille, S., Battesti, E., Filliat, D., 2015. A light visual mapping and navigation framework for low-cost robots. J. Intell. Syst. 24 (4), 505–524.

Betancourt, A., Morerio, P., Regazzoni, C., Rauterberg, M., 2015. The evolution of first person vision methods: asurvey. Circ. Syst. Video Technol. IEEE Trans. 25 (5), 744–760.

Bishop, J., 2014. Supporting communication between people with social orientation impairments using affective computing technologies. In: Assistive Technologies for Physical and Cognitive Disabilities, p. 42.

Bonnet, V., Coste, C.A., Lapierre, L., Cadic, J., Fraisse, P., Zapata, R., Venture, G., Geny, C., 2015. Towards an affordable mobile analysis platform for pathological walking assessment. Robot. Auton. Syst. 66, 116–128. doi:10.1016/j.robot.2014.12.002.

Bosch, N., DMello, S., Baker, R., Ocumpaugh, J., Shute, V., Ventura, M., Wang, L., Zhao, W., 2015. Automatic detection of learning-centered affective states in the wild. In: Proceedings of the 2015 International Conference on Intelligent User Interfaces (IUI 2015). ACM, New York, NY, USA.

Bourbakis, N., Papadakis Ktistakis, I., Tsoukalas, L., Alamaniotis, M., 2015. An autonomous intelligent wheelchair for assisting people at need in smart homes: A case study. In: International Conference on Information, Intelligence, Systems and Applications, July 6–8, 2015, Corfu, Greece. IEEE, pp. 1–7.

Bowes, A., Dawson, A., Greasley-Adams, C., 2013. Literature review: the cost effectiveness of assistive technology in supporting people with dementia. Published by the Dementia Services Development Centre, University of Stirling.

Bruna, A., Farinella, G.M., Guarnera, G.C., Battiato, S., 2013. Forgery detection and value identification of euro banknotes. Sensors 13 (2), 2515–2529.

BrgidaMnica, F., LusPaulo, R., Nuno, L., 2015. A methodology for creating an adapted command language for driving an intelligent wheelchair. J. Intell. Robot. Syst. 1–15.

Canal, G., Escalera, S., Angulo, C., 2016. A real-time human-robot interaction system based on gestures for assistive scenarios. Comput. Vision Image Understand. 149, 65–77.

Carcagnì, P., Cazzato, D., Del Coco, M., Distante, C., Leo, M., 2015. Visual interaction including biometrics information for a socially assistive robotic platform. In: Agapito, L., Bronstein, M.M., Rother, C. (Eds.), Computer Vision - ECCV 2014 Workshops. In: Lecture Notes in Computer Science, 8927. Springer International Publishing, pp. 391–406.

Carcagnì, P., Cazzato, D., Del Coco, M., Mazzeo, P.L., Leo, M., Distante, C., 2015. Soft biometrics for a socially assistive robotic platform. Paladyn J. Behav. Robot. 6 (1).

Cassidy, S., Stenger, B., Van Dongen, L., Yanagisawa, K., Anderson, R., Wan, V., Baron–Cohen, S., Cipolla, R., 2016. Expressive visual text-to-speech as an assistive technology for individuals with autism spectrum conditions. Comput. Vision Image Understand. 148, 193–200.

Cazzato, D., Leo, M., Distante, C., 2014. An investigation on the feasibility of uncalibrated and unconstrained gaze tracking for human assistive applications by using head pose estimation. Sensors 14 (5), 8363–8379.

Chang, H.J., Garcia-Hernando, G., Tang, D., Kim, T.-K., 2016. Spatio-temporal hough forest for efficient detectionlocalisationrecognition of fingerwriting in egocentric camera. Comput. Vision Image Understand. 148, 87–96.

Chaquet, J.M., Carmona, E.J., Fernndez-Caballero, A., 2013. A survey of video datasets for human action and activity recognition. Comput. Vision Image Understand. 117 (6), 633–659.

Chaudhry, S., Chandra, R., 2015. Design of a mobile face recognition system for visually impaired persons. arXiv preprint arXiv:1502.00756.

Chen, S., Liu, C., 2015. Eye detection using discriminatory haar features and a new efficient svm. Image Vision Comput. 33, 68–77.

Chessa, M., Noceti, N., Odone, F., Solari, F., Sosa-García, J., Zini, L., 2016. An integrated artificial vision framework for assisting visually impaired users. Comput. Vision Image Understand. 149, 209–228.

Chippendale, P., Tomaselli, V., D'Alto, V., Urlini, G., Modena, C.M., Messelodi, S., Strano, S.M., Alce, G., Hermodsson, K., Razafimahazo, M., et al., 2014. Personal shopping assistance and navigator system for visually impaired people. ACVR2014: Second Workshop on Assistive Computer Vision and Robotics.

Cook, A.M., Polgar, J.M., 2014. Assistive Technologies: Principles and Practice. Elsevier Health Sciences.

Cooper, R.A., 2007. An Introduction to Rehabilitation Engineering. CRC Press.

Cormier, M., Moffatt, K., Cohen, R., Mann, R., 2016. Purely vision-based segmentation of web pages for assistive technology. Comput. Vision Image Understand. 148, 46–66.

Coughlan, J., Manduchi, R., 2009. Functional assessment of a camera phone-based wayfinding system operated by blind and visually impaired users. International Journal on Artificial Intelligence Tools 18 (03), 379–397.

Coughlan, J.M., Shen, H., 2013. Crosswatch: a system for providing guidance to visually impaired travelers at traffic intersection. J. Assistive Technol. 7 (2), 131–142.

Cowan, R.E., Fregly, B.J., Boninger, M.L., Chan, L., Rodgers, M.M., Reinkensmeyer, D.J., et al., 2012. Recent trends in assistive technology for mobility. J. Neuroeng. Rehab. 9 (1), 20.

Cristani, M., Murino, V., 2014. Socially-driven computer vision for group behavior analysis. In: Registration and Recognition in Images and Videos. In: Studies in Computational Intelligence, 532. Springer Berlin Heidelberg, pp. 223–256.

Cristina, S., Camilleri, K.P., 2016. Model-based head pose-free gaze estimation for assistive communication. Comput. Vision Image Understand. 149, 157–170.

Dakopoulos, D., Bourbakis, N., 2010. Wearable obstacle avoidance electronic travel aids for blind: a survey. IEEE Transactions on Syst. Man Cybern. Part C 40 (1), 25–35.

Damen, D., Leelasawassuk, T., Haines, O., Calway, A., Mayol-Cuevas, W., 2014. You-do, i-learn: discovering task relevant objects and their modes of interaction from multi-user egocentric video. In: British Machine Vision Conference.

Damen, D., Leelasawassuk, T., Mayol-Cuevas, W., 2016. You-do, i-learn: egocentric unsupervised discovery of objects and their modes of interaction towards video-based guidance. Comput. Vision Image Understand. 149, 98–112.

Delibasis, K.K., Plagianakos, V.P., Maglogiannis, I., 2014. Refinement of human silhouette segmentation in omni-directional indoor videos. Comput. Vision Image Understand. 128, 65–83.

DemCare, 2016. Dementia ambient care: multi-sensing monitoring for intelligent remote management and decision support. URL http://www.demcare.eu/.

Devyver, M., Tsukada, A., Kanade, T., 2011. A wearable device for first person vision. International Symposium on Quality of Life Technology.

Devyver, M.S., Tsukada, A., Kanade, T., 2011. A wearable device for first person vision. 3rd International Symposium on Quality of Life Technology.

Di Lascio, R., Foggia, P., Percannella, G., Saggese, A., Vento, M., 2013. A real time algorithm for people tracking using contextual reasoning. Comput. Vision Image Understand. 117 (8), 892–908.

Diraco, G., Leone, A., Siciliano, P., 2015. People occupancy detection and profiling with 3d depth sensors for building energy management. Energy Build. 92, 246–266.

Dollar, P., Appel, R., Belongie, S., Perona, P., 2014. Fast feature pyramids for object detection. Pattern Anal. Mach. Intell. IEEE Trans. 36 (8), 1532–1545.

Donahue, J., Anne Hendricks, L., Guadarrama, S., Rohrbach, M., Venugopalan, S., Saenko, K., Darrell, T., 2015. Long-term recurrent convolutional networks for visual recognition and description. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2625–2634.

Došen, S., Cipriani, C., Kostić, M., Controzzi, M., Carrozza, M.C., Popović, D.B., 2010. Cognitive vision system for control of dexterous prosthetic hands: experimental evaluation. J. Neuroeng. Rehab. 7 (1), 42.

Endres, F., Hess, J., Engelhard, N., Sturm, J., Cremers, D., Burgard, W., 2012. An evaluation of the RGB-D SLAM system. In: Robotics and Automation (ICRA), 2012 IEEE International Conference on. IEEE, pp. 1691–1696.

Engelhard, N., Endres, F., Hess, J., Sturm, J., Burgard, W., 2011. Real-time 3d visual SLAM with a hand-held RGB-Dcamera. In: Proc. of the RGB-D Workshop on 3D Perception in Robotics at the European Robotics Forum, Vasteras, Sweden, 180.

Epshtein, B., Ofek, E., Wexler, Y., 2010. Detecting text in natural scenes with stroke width transform. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2963–2970.

Escobedo, L., Tentori, M., Quintana, E., Favela, J., Garcia-Rosas, D., 2014. Using augmented reality to help children with autism stay focused. Pervasive Comput. IEEE 13 (1), 38–46.

Eweiwi, A., Cheema, M.S., Bauckhage, C., 2015. Action recognition in still images by learning spatial interest regions from videos. Pattern Recognit. Lett. 51, 8–15. doi:10.1016/j.patrec.2014.07.017.

Fahad, L.G., Khan, A., Rajarajan, M., 2015. Activity recognition in smart homes with self verification of assignments. Neurocomputing 149, 1286–1298.

Fallon, M.F., Johannsson, H., Leonard, J.J., 2012. Efficient scene simulation for robust monte carlo localization using an RGB-D camera. In: Robotics and Automation (ICRA), 2012 IEEE International Conference on. IEEE, pp. 1663–1670.

Farinella, G.M., Allegra, D., Moltisanti, M., Stanco, F., Battiato, S., 2016. Retrieval and classification of food images. Comput. Biol. Med. 77, 23–39.

Fasola, J., Mataric, M., 2013. A socially assistive robot exercise coach for the elderly. J. Hum. Rob. Interact. 2 (2), 3–32.

Federici, S., Scherer, M., 2012. Assistive Technology Assessment Handbook. CRC Press.

Feng, W., Liu, R., Zhu, M., 2014. Fall detection for elderly person care in a vision-based home surveillance environment using a monocular camera. Signal Image Video Process. 8 (6), 1129–1138.

Fogelton, A., Benesova, W., 2016. Eye blink detection based on motion vectors analysis. Comput. Vision Image Understand. 148, 23–33.

Forkan, A.R.M., Khalil, I., Tari, Z., Foufou, S., Bouras, A., 2015. A context-aware approach for long-term behavioural change detection and abnormality prediction in ambient assisted living. Pattern Recognit. 48 (3), 628–641.

Fuhrer, M., Jutai, J., Scherer, M., DeRuyter, F., 2003. A framework for the conceptual modelling of assistive technology device outcomes. Disability Rehab. 25 (22), 1243–1251.

Furnari, A., Farinella, G.M., Battiato, S., 2015. Recognizing personal contexts from egocentric images. In: International Workshop on Assistive Computer Vision and Robotics - IEEE International Conference on Computer Vision Workshop (ICCVW), pp. 393–401.

Furnari, A., Farinella, G.M., Battiato, S., 2016. Temporal segmentation of egocentric videos to highlight personal locations of interest. In: International Workshop on Egocentric Perception, Interacion, and Computing - European Conference on Computer Vision Workshop.

Gamallo, C., Mucientes, M., Regueiro, C.V., 2015. Omnidirectional visual slam under severe occlusions. Robot. Auton. Syst. 65, 76–87.

Giakoumis, D., Stavropoulos, G., Kikidis, D., Vasileiadis, M., Votis, K., Tzovaras, D., 2014. Recognizing daily activities in realistic environments through depth-based user tracking and hidden conditional random fields for MCI/AD support. In: Computer Vision-ECCV 2014 Workshops, pp. 822–838.

Gillespie, A., Best, C., O'Neill, B., 2012. Cognitive function and assistive technology for cognition: a systematic review. J. Int. Neuropsychol. Soc. 18 (01), 1–19.

Gkioxari, G., Malik, J., 2015. Finding action tubes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 759–768.

Gurrin, C., Smeaton, A.F., Doherty, A.R., 2014. Lifelogging: personal big data. Found. Trends Inf. Retrieval 8 (1), 1–125.

Haltakov, V., Unger, C., Ilic, S., 2016. Geodesic pixel neighborhoods for 2d and 3d scene understanding. Comput. Vision Image Understand. 148, 164–180.

Hersh, M., Johnson, M.A., 2010. Assistive Technology for Visually Impaired and Blind People. Springer Science & Business Media.

Ho, E.S., Chan, L.C., Chan, D.C., Shum, H.P., Cheung, Y.-M., Yuen, P.C., 2016. Improving posture classification accuracy for depth sensor-based human activity monitoring in smart environments. Comput. Vision Image Understand. 148, 97–110.

Horne, L., Alvarez, J., McCarthy, C., Salzmann, M., Barnes, N., 2016. Semantic labeling for prosthetic vision. Comput. Vision Image Understand. 149, 113–125.

Geneva: World Health Organization Concept Note: Opening the GATE for Assistive Health Technology: Shifting the paradigm. 2015. WHO, http://www.who.int/.

Huang, G., Pun, C.-M., Lin, C., Zhou, Y., 2015. Non-rigid visual object tracking using user-defined marker and gaussian kernel. Multimedia Tools Appl. 1–20.

Iscen, A., Wang, Y., Duygulu, P., Hauptmann, A., 2015. Snippet based trajectory statistics histograms for assistive technologies. In: Agapito, L., Bronstein, M.M., Rother, C. (Eds.), Computer Vision - ECCV 2014 Workshops. In: Lecture Notes in Computer Science, 8928. Springer International Publishing, pp. 3–16.

Jaderberg, M., Simonyan, K., Vedaldi, A., Zisserman, A., 2014. Reading text in the wild with convolutional neural networks. arXiv preprint arXiv:1412.1842.

Jafri, R., Ali, S., 2014. A multimodal tabletbased application for the visually impaired for detecting and recognizing objects in a home environment. In: Miesenberger, K., Fels, D., Archambault, D., Pez, P., Zagler, W. (Eds.), Computers Helping People with Special Needs. In: Lecture Notes in Computer Science, 8547. Springer International Publishing, pp. 356–359.

Jafri, R., Ali, S.A., Arabnia, H.R., Fatima, S., 2014. Computer vision-based object recognition for the visually impaired in an indoors environment: a survey. Visual Comput. 30 (11), 1197–1222.

Jeni, L.A., Cohn, J.F., Kanade, T., 2015. Dense 3d face alignment from 2d videos in real-time. In: 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG).

Jiang, H., Zhang, T., Wachs, J.P., Duerstock, B.S., 2016. Enhanced control of a wheelchair-mounted robotic manipulator using 3-d vision and multimodal interaction. Comput. Vision Image Understand. 149, 21–31.

Jo, H.S., Tsun, M.T.K., Theng, L.B., Hui, P.T.H., 2014. Robotics for assisting children with physical and cognitive disabilities. Assistive Technol. Phys. Cognit. Disabilities 78.

Jo, J., Choi, H., Kim, I.-J., Kim, J., 2015. Single-view-based 3d facial reconstruction method robust against pose variations. Pattern Recognit. 48 (1), 73–85.

Jutai, J.W., Fuhrer, M.J., Demers, L., Scherer, M.J., DeRuyter, F., 2005. Toward a taxonomy of assistive technology device outcomes. Am. J. Phys. Med. Rehab. 84 (4), 294–302.

Kanade, T., Hebert, M., 2012. First-person vision. Proc. IEEE 100 (8), 2442–2453.

Kang, H., Hebert, M., Efros, A., Kanade, T., 2015. Data-driven objectness. Pattern Anal. Mach. Intell. IEEE Trans. 37 (1), 189–195.

Keefer, R., Bourbakis, N., 2014. From image to XML: monitoring a page layout analysis approach for the visually impaired. Int. J. Monit. Surveillance Technol. Res. 2 (1), 22–43.

Keefer, R., Liu, Y., Bourbakis, N., 2013. The development and evaluation of an eye-s-free interaction model for mobile reading devices. IEEE Trans. Hum. Mach. Syst. 43 (1), 76–91.

Kendall, A., Grimes, M., Cipolla, R., 2015. Posenet: a convolutional network for real-time 6-DOFcamera relocalization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2938–2946.

Kerl, C., Sturm, J., Cremers, D., 2013. Dense visual SLAM for RGB-D cameras. In: Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on. IEEE, pp. 2100–2106.

Koo, H.I., Cho, N.I., 2010. State estimation in a document image and its application in text block identification and text line extraction. In: European Conference on Computer Vision. Springer, pp. 421–434.

Lahav, O., Schloerb, D.W., Srinivasan, M.A., 2015. Rehabilitation program integrating virtual environment to improve orientation and mobility skills for people who are blind. Comput. Edu. 80, 1–14.

Lancioni, G.E., Sigafoos, J., O'Reilly, M.F., Singh, N.N., 2012. Assistive Technology: Interventions for Individuals with Severe/Profound and Multiple Disabilities. Springer Science & Business Media.

Lancioni, G.E., Singh, N.N., 2014. Assistive Technologies for People with Diverse Abilities. Springer.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.

Lee, Y., Medioni, G., 2015. Wearable RGBD indoor navigation system for the blind. In: Agapito, L., Bronstein, M.M., Rother, C. (Eds.), Computer Vision - ECCV 2014 Workshops. In: Lecture Notes in Computer Science, 8927. Springer International Publishing, pp. 493–508.

Lee, Y.H., Medioni, G., 2016. RGB-D camera based wearable navigation system for the visually impaired. Comput. Vision Image Understand. 149, 3–20.

Lee, Y.J., Ghosh, J., Grauman, K., 2012. Discovering important people and objects for egocentric video summarization. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1346–1353.

Lee, Y.J., Grauman, K., 2015. Predicting important objects for egocentric video summarization. Int. J. Comput. Vision 114 (1).

Lenker, J.A., Harris, F., Taugher, M., Smith, R.O., 2013. Consumer perspectives on assistive technology outcomes. Disability Rehab. 8 (5), 373–380. doi:10.3109/17483107.2012.749429.

Leo, M., Cazzato, D., De Marco, T., Distante, C., 2014. Unsupervised eye pupil localization through differential geometry and local self-similarity matching. PloS One 9 (8), e102829.

Leung, T.-S., Medioni, G., 2014. Visual navigation aid for the blind in dynamic environments. In: Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on. IEEE, pp. 579–586.

Li, K., Fu, Y., 2014. Prediction of human activity by discovering temporal sequence patterns. Pattern Anal. Mach. Intell. IEEE Trans. 36 (8), 1644–1657.

Li, L.-J., Su, H., Lim, Y., Fei-Fei, L., 2014. Object bank: an object-level image representation for high-level visual recognition. Int. J. Comput. Vision 107 (1), 20–39.

Li, Y.-T., Wachs, J.P., 2014. HEGM: a hierarchical elastic graph matching for hand gesture recognition. Pattern Recognit. 47 (1), 80–88.

Liu, X., Cheung, Y.-M., Tang, Y.Y., 2016. Lip event detection using oriented histograms of regional optical flow and low rank affinity pursuit. Comput. Vision Image Understand. 148, 153–163.

Louie, W.-Y.G., McColl, D., Nejat, G., 2014. Acceptance and attitudes toward a human-like socially assistive robot by older adults. Assistive Technol. 26 (3), 140–150.

Lu, Z., Grauman, K., 2013. Story-driven summarization for egocentric video. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2714–2721.

Luo, J., Wang, W., Qi, H., 2014. Spatio-temporal feature extraction and representation for RGB-D human action recognition. Pattern Recognit. Lett. 50, 139–148.

Ma, B., Huang, R., Qin, L., 2015. VoD: a novel image representation for head yaw estimation. Neurocomputing 148, 455–466.

Madadi, M., Escalera, S., Gonzàlez, J., Roca, F.X., Lumbreras, F., 2015. Multi-part body segmentation based on depth maps for soft biometry analysis. Pattern Recognit. Lett. 56, 14–21.

Mann, S., 1997. Wearable computing: a first step toward personal imaging. Computer 30 (2), 25–32.

Marcon, M., Sarti, A., Tubaro, S., 2016. Toothbrush motion analysis to help children learn proper tooth brushing. Comput. Vision Image Understand. 148, 34–45.

Markuš, N., Frljak, M., Pandžić, I.S., Ahlberg, J., Forchheimer, R., 2014. Eye pupil localization with an ensemble of randomized trees. Pattern Recognit. 47 (2), 578–587.

Martinel, N., Piciarelli, C., Micheloni, C., 2016. A supervised extreme learning committee for food recognition. Comput. Vision Image Understand. 148, 67–86.

Mascetti, S., Ahmetovic, D., Gerino, A., Bernareggi, C., Busso, M., Rizzi, A., 2016. Robust traffic lights detection on mobile devices for pedestrians with visual impairment. Comput. Vision Image Understand. 148, 123–135.

Mataric, M.J., 1999. Socially assistive robotics. Science 3 (6), 233–242.

Matsuo, K., Yamada, K., Ueno, S., Naito, S., 2014. An attention-based activity recognition for egocentric video. In: Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on, pp. 565–570.
Matusiak, K., Skulimowski, P., Strumillo, P., 2014. A mobile phone application for recognizing objects as a personal aid for the visually impaired users. In: Hippe, Z.S., Kulikowski, J.L., Mroczek, T., Wtorek, J. (Eds.), Human-Computer Systems Interaction: Backgrounds and Applications 3. In: Advances in Intelligent Systems and Computing, 300. Springer International Publishing, pp. 201–212.
Mead, R., Atrash, A., Matarić, M.J., 2013. Automated proxemic feature extraction and behavior recognition: Applications in human-robot interaction. Int. J. Social Robot. 5 (3), 367–378.
Medioni, G., François, A.R., Siddiqui, M., Kim, K., Yoon, H., 2007. Robust real-time vision for a personal service robot. Comput. Vision Image Understand. 108 (1), 196–203.
Memon, M., Wagner, S.R., Pedersen, C.F., Beevi, F.H.A., Hansen, F.O., 2014. Ambient assisted living healthcare frameworks, platforms, standards, and quality attributes. Sensors 14 (3), 4312–4341.
Meng, L., De Silva, C., Zhang, J., 2014. 3d visual SLAM for an assistive robot in indoor environments using RGB-D cameras. In: Computer Science Education (ICCSE), 2014 9th International Conference on, pp. 32–37.
Mihailidis, A., Carmichael, B., Boger, J., 2004. The use of computer vision in an intelligent environment to support aging-in-place, safety, and independence in the home. Inf. Technol. Biomed. IEEE Trans. 8 (3), 238–247.
Molina, E., Diallo, A., Zhu, Z., 2013. Visual noun navigation framework for the blind. J. Assistive Technol. 7 (2), 118–130.
Mollaret, C., Mekonnen, A.A., Lerasle, F., Ferrané, I., Pinquier, J., Boudet, B., Rumeau, P., 2016. A multi-modal perception based assistive robotic system for the elderly. Comput. Vision Image Understand. 149, 78–97.
Mora, K.A.F., Odobez, J.-M., 2014. Geometric generative gaze estimation (g3e) for remote RGB-D cameras. In: Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. IEEE, pp. 1773–1780.
Mubashir, M., Shao, L., Seed, L., 2013. A survey on fall detection: principles and approaches. Neurocomputing 100, 144–152.
Murphy, K., Darrah, M., 2015. Haptics-based apps for middle school students with visual impairments. IEEE Trans. Haptics 8 (3), 318–326.
Narayanan, V.K., Pasteau, F., Babel, M., Chaumette, F., 2014. Visual servoing for autonomous doorway passing in a wheelchair using a single doorpost. IEEE/RSJ IROS Workshop on Assistance and Service Robotics in a Human Environment, ASROB.
Narayanan, V.K., Pasteau, F., Marchal, M., Krupa, A., Babel, M., 2016. Vision-based adaptive assistance and haptic guidance for safe wheelchair corridor following. Comput. Vision Image Understand. 149, 171–185.
Nguyen, Q.-H., Vu, H., Tran, T.-H., Van Hamme, D., Veelaert, P., Philips, W., Nguyen, Q.-H., 2015. A visual SLAM system on mobile robot supporting localization services to visually impaired people. In: Agapito, L., Bronstein, M.M., Rother, C. (Eds.), Computer Vision - ECCV 2014 Workshops. Springer International Publishing, pp. 716–729.
Ohn-Bar, E., Trivedi, M., 2014. Hand gesture recognition in real time for automotive interfaces: a multimodal vision-based approach and evaluations. Intell. Transp. Syst. IEEE Trans. 15 (6), 2368–2377.
Okeyo, G., Chen, L., Wang, H., Sterritt, R., 2014. Dynamic sensor data segmentation for real-time knowledge-driven activity recognition. Pervasive Mobile Comput. 10 (Part B), 155–172.
Oliver, M., Montero, F., Fernndez-Caballero, A., Gonzlez, P., Molina, J.P., 2015. RGB-D assistive technologies for acquired brain injury: description and assessment of user experience. Expert Syst. 32 (3), 370–380.
OrCam, 2016. MyeyeURL http://www.orcam.com/.
Ortis, A., Farinella, G.M., D'Amico, V., Addesso, L., Torrisi, G., Battiato, S., 2016. Organizing egocentric videos for daily living monitoring. Lifelogging Tools and Applications - ACM MM Workshop.
O'Neill, B., Gillespie, A., 2014. Assistive technology for cognition. In: Assistive Technology for Cognition: A Handbook for Clinicians and Developers. Psychology Press.
Panteleris, P., Argyros, A.A., 2014. Vision-based SLAM and moving objects tracking for the perceptual support of a smart walker platform. In: Computer Vision-ECCV 2014 Workshops. Springer, pp. 407–423.
Papoutsakis, K., Padeleris, P., Ntelidakis, A., Stefanou, S., Zabulis, X., Kosmopoulos, D., Argyros, A.A., 2013. Developing visual competencies for socially assistive robots: the hobbit approach. In: Proceedings of the 6th International Conference on PErvasive Technologies Related to Assistive Environments. ACM, p. 56.
Parkhi, O.M., Vedaldi, A., Zisserman, A., 2015. Deep face recognition. In: British Machine Vision Conference, vol. 1, p. 6.
Pasteau, F., Krupa, A., Babel, M., 2014. Vision-based assistance for wheelchair navigation along corridors. In: Robotics and Automation (ICRA), 2014 IEEE International Conference on. IEEE, pp. 4430–4435.
Pawluk, D., Adams, R., Kitada, R., 2015. Designing haptic assistive technology for individuals who are blind or visually impaired. IEEE Trans. Haptics 8 (3), 245–257.
Pawluk, D., Bourbakis, N., Giudice, N., Hayward, V., Heller, M., 2015. Guest editorial: Haptic assistive technology for individuals who are visually impaired. IEEE Trans. Haptics 8 (3), 245–257.
Phung, S.L., Le, M.C., Bouzerdoum, A., 2016. Pedestrian lane detection in unstructured scenes for assistive navigation. Comput. Vision Image Understand. 149, 186–196.
Pirsiavash, H., Ramanan, D., 2012. Detecting activities of daily living in first-person camera views. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pp. 2847–2854.
Piyathilaka, L., Kodagoda, S., 2015. Human activity recognition for domestic robots. In: Field and Service Robotics. Springer, pp. 395–408.
Poggi, M., Nanni, L., Mattoccia, S., 2015. Crosswalk recognition through point-cloud processing and deep-learning suited to a wearable mobility aid for the visually impaired. In: International Conference on Image Analysis and Processing. Springer, pp. 282–289.
Poleg, Y., Arora, C., Peleg, S., 2014. Temporal segmentation of egocentric videos. In: IEEE Conference on Computer Vision and Pattern Recognition.
Rabbitt, S.M., Kazdin, A.E., Scassellati, B., 2015. Integrating socially assistive robotics into mental healthcare interventions: Applications and recommendations for expanded use. Clin. Psychol. Rev. 35, 35–46.
Rahmani, H., Mian, A., Shah, M., 2016. Learning a deep model for human action recognition from novel viewpoints. arXiv preprint arXiv:1602.00828.
Rautaray, S., Agrawal, A., 2015. Vision based hand gesture recognition for human computer interaction: a survey. Artif. Intell. Rev. 43 (1), 1–54.
Rehg, J.M., Rozga, A., Abowd, G.D., Goodwin, M.S., 2014. Behavioral imaging and autism. Pervasive Comput. IEEE 13 (2), 84–87.
Rivera, L.A., DeSouza, G., Franklin, L., 2013. Control of a wheelchair using an adaptive k-means clustering of head poses. In: Computational Intelligence in Rehabilitation and Assistive Technologies (CIRAT), 2013 IEEE Symposium on. IEEE, pp. 24–31.
Rivera-Rubio, J., Alexiou, I., Bharath, A.A., 2015. Appearance-based indoor localization: a comparison of patch descriptor performance. Pattern Recognit. Lett. 66, 109–117.
Rivera-Rubio, J., Arulkumaran, K., Rishi, H., Alexiou, I., Bharath, A.A., 2016. An assistive haptic interface for appearance-based indoor navigation. Comput. Vision Image Understand. 149, 126–145.
Rowekamper, J., Sprunk, C., Tipaldi, G.D., Stachniss, C., Pfaff, P., Burgard, W., 2012. On the position accuracy of mobile robot localization based on particle filters combined with scan matching. In: Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on. IEEE, pp. 3158–3164.
Raví, D., Bober, M., Farinella, G., Guarnera, M., Battiato, S., 2016. Semantic segmentation of images exploiting DCT based features and random forest. Pattern Recognit. 52, 260–273.
Ryoo, M., Fuchs, T.J., Xia, L., Aggarwal, J., Matthies, L., 2015. Robot-centric activity prediction from first-person videos: what will they do to me'. In: Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction. ACM, pp. 295–302.
Salah, A.A., Sebe, N., Gevers, T., 2010. Communication and automatic interpretation of affect from facial expressions. In: Affective Computing and Interaction: Psychological, Cognitive and Neuroscientific Perspectives: Psychological, Cognitive and Neuroscientific Perspectives, p. 157.
Sánchez, C., Taddei, P., Ceriani, S., Wolfart, E., Sequeira, V., 2016. Localization and tracking in known large environments using portable real-time 3d sensors. Comput. Vision Image Understand. 149, 197–208.
Sandygulova, A., Dragone, M., O'Hare, G., 2014. Real-time adaptive child-robot interaction: age and gender determination of children based on 3d body metrics. In: Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on, pp. 826–831.
Schmidt, A., Kraft, M., 2015. The impact of the image feature detector and descriptor choice on visual SLAM accuracy. In: Image Processing & Communications Challenges 6. Springer, pp. 203–210.
Sefidgar, Y.S., Vahdat, A., Se, S., Mori, G., 2015. Discriminative key-component models for interaction detection and recognition. Comput. Vision Image Understand. 135, 16–30.
Seo, J., Han, S., Lee, S., Kim, H., 2015. Computer vision techniques for construction safety and health monitoring. Adv. Eng. Inf 29 (2), 239–251.
Sheikhi, S., Odobez, J.-M., 2015. Combining dynamic head posegaze mapping with the robot conversational state for attention recognition in humanrobot interactions. Pattern Recognit. Lett. 66, 81–90. Pattern Recognition in Human Computer Interaction. doi: 10.1016/j.patrec.2014.10.002.
Shinohara, K., Wobbrock, J.O., 2016. Self-conscious or self-confident? a diary study conceptualizing the social accessibility of assistive technology. ACM Trans. Accessible Comput. 8 (2), 5.
Shoaib, M., Bosch, S., Incel, O.D., Scholten, H., Havinga, P.J., 2015. A survey of online activity recognition using mobile phones. Sensors 15 (1), 2059–2085.
Sivalingam, R., Cherian, A., Fasching, J., Walczak, N., Bird, N., Morellas, V., Murphy, B., Cullen, K., Lim, K., Sapiro, G., et al., 2012. A multi-sensor visual tracking system for behavior monitoring of at-risk children. In: Robotics and Automation (ICRA), 2012 IEEE International Conference on. IEEE, pp. 1345–1350.
Sivaraman, S., Trivedi, M., 2014. Active learning for on-road vehicle detection: a comparative study. Mach. Vision Appl. 25 (3), 599–611.
Sobia, M., Brindha, V., Abudhahir, A., 2014. Facial expression recognition using PCA based interface for wheelchair. In: Electronics and Communication Systems (ICECS), 2014 International Conference on, pp. 1–6.
Sun, M., Savarese, S., 2014. Model-based object recognition. In: Ikeuchi, K. (Ed.), Computer Vision. Springer US, pp. 488–492.
Takizawa, H., Yamaguchi, S., Aoyagi, M., Ezaki, N., Mizuno, S., 2015. Kinect cane: an assistive system for the visually impaired based on the concept of object recognition aid. Pers. Ubiquitous Comput. 1–11.
Tao, L., Paiement, A., Damen, D., Mirmehdi, M., Hannuna, S., Camplani, M., Burghardt, T., Craddock, I., 2016. A comparative study of pose representation and dynamics modelling for online motion quality assessment. Comput. Vision Image Understand. 148, 136–152.

Thakoor, K., Mante, N., Zhang, C., Siagian, C., Weiland, J., Itti, L., Medioni, G., 2015. A system for assisting the visually impaired in localization and grasp of desired objects. In: Agapito, L., Bronstein, M.M., Rother, C. (Eds.), Computer Vision - ECCV 2014 Workshops. In: Lecture Notes in Computer Science, vol. 8927. Springer International Publishing, pp. 643–657.

Tian, Y., Yang, X., Yi, C., Arditi, A., 2013. Toward a computer vision-based wayfinding aid for blind persons to access unfamiliar indoor environments. Mach. Vision Appl. 24 (3), 521–535.

Tong, Y., Chen, R., Gao, J., 2015. Hidden state conditional random field for abnormal activity recognition in smart homes. Entropy 17 (3), 1358–1378.

Topal, C., Gunal, S., Koçdeviren, O., Dogan, A., Gerek, O.N., 2014. A low-computational approach on gaze estimation with eye touch system. Cybern. IEEE Trans. 44 (2), 228–239.

Tsoukalas, A., Bourbakis, N., 2013. A first stage comparative survey on human activity recognition methodologies. Int. J. Artif. Intell. Tools 24 (6).

Unar, J., Seng, W.C., Abbasi, A., 2014. A review of biometric technology along with trends and prospects. Pattern Recognit. 47 (8), 2673–2688.

Velázquez, R., 2010. Wearable assistive devices for the blind. In: Wearable and Autonomous Biomedical Devices and Systems for Smart Environment. Springer, pp. 331–349.

Vichitvanichphong, S., Talaei-Khoei, A., Kerr, D., Ghapanchi, A.H., 2014. Adoption of assistive technologies for aged care: arealist review of recent studies. In: System Sciences (HICSS), 2014 47th Hawaii International Conference on. IEEE, pp. 2706–2715.

Villamizar, M., Garrell, A., Sanfeliu, A., Moreno-Noguer, F., 2016. Interactive multiple object learning with scanty human supervision. Comput. Vision Image Understand. 149, 51–64.

Vuong, N.K., Chan, S., Lau, C.T., 2015. mhealth sensors, techniques, and applications for managing wandering behavior of people with dementia: a review. In: Mobile Health. Springer, pp. 11–42.

Wang, J., Liu, Z., Wu, Y., Yuan, J., 2014. Learning actionlet ensemble for 3d human action recognition. Pattern Anal. Mach. Intell. IEEE Trans. 36 (5), 914–927.

Wang, J., Xiong, R., Chu, J., 2015. Facial feature points detecting based on gaussian mixture models. Pattern Recognit. Lett. 53, 62–68. doi:10.1016/j.patrec.2014.11.004.

Wang, P., Ma, S., Shen, Y., 2014. Performance study of feature descriptors for human detection on depth map. Int. J. Model. Simul. Sci. Comput. 5 (03).

Wang, P., Sun, L., Yang, S., Smeaton, A.F., Gurrin, C., 2016. Characterizing everyday activities from visual lifelogs based on enhancing concept representation. Comput. Vision Image Understand. 148, 181–192.

Wei, X., Phung, S.L., Bouzerdoum, A., 2014. Object segmentation and classification using 3-d range camera. J. Visual Commun. Image Represent. 25 (1), 74–85.

Wilkowska, W., Gaul, S., Ziefle, M., 2010. A Small but Significant Difference–The Role of Gender on Acceptance of Medical Assistive Technologies. Springer.

Wu, Y., Ma, B., Jia, Y., 2015. Differential tracking with a kernel-based region covariance descriptor. Pattern Anal. Appl. 18 (1), 45–59.

Yan, W., Weber, C., Wermter, S., 2011. A hybrid probabilistic neural model for person tracking based on a ceiling-mounted camera. J. Ambient Intell. Smart Environ. 3 (3), 237–252.

Yu, L., Ong, S.K., Nee, A.Y.C., 2016. A tracking solution for mobile augmented reality based on sensor-aided marker-less tracking and panoramic mapping. Multimedia Tools Appl. 75 (6), 3199–3220.

Yun, Y., Gu, I.Y.-H., 2016. Human fall detection in videos via boosting and fusing statistical features of appearance, shape and motion dynamics on riemannian manifolds with applications to assisted living. Comput. Vision Image Understand. 148, 111–122.

Zabala, J., 1995. The sett framework: critical areas to consider when making informed assistive technology decisions.

Zagoruyko, S., Lerer, A., Lin, T.-Y., Pinheiro, P. O., Gross, S., Chintala, S., Dollár, P., 2016. A multipath network for object detection. arXiv preprintarXiv:1604.02135.

Zeng, J., Sun, Y., Wang, F., 2012. A natural hand gesture system for intelligent human-computer interaction and medical assistance. In: Intelligent Systems (GCIS), 2012 Third Global Congress on, pp. 382–385.

Zhai, M., Roshtkhari, M. J., Mori, G., 2016. Deep learning of appearance models for online object tracking. arXiv preprint arXiv:1607.02568.

Zhang, H., Beveridge, J.R., Draper, B.A., Phillips, P.J., 2015. On the effectiveness of soft biometrics for increasing face verification rates. Comput. Vision Image Understand. 137, 50–62.

Zhang, J., Shan, Y., Huang, K., 2015. ISEE smart home (ISH): Smart video analysis for home security. Neurocomputing 149, 752–766.

Zhang, W., Smith, M.L., Smith, L.N., Farooq, A., 2016. Gender and gaze gesture recognition for human-computer interaction. Comput. Vision Image Understand. 149, 32–50.

Ziefle, M., Klack, L., Wilkowska, W., Holzinger, A., 2013. Acceptance of telemedical treatments–a medical professional point of view. In: Human Interface and the Management of Information. Information and Interaction for Health, Safety, Mobility and Complex Environments. Springer, pp. 325–334.